

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 04:55:59 ; Search time 2988 Seconds
(without alignments)
18293.095 Million cell updates/sec

Title: US-09-930-020A-1

Perfect score: 3375

Sequence: 1 gacagtgttcggtgcac.....tgtccacctgaaggtcttc 3375

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_estc:*
- 9: gb_estl:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	305	9.0	535	12	BF849816
C 2	296	8.8	406	10	AW206492
C 3	128	3.8	282	10	BE143292
C 4	113	3.3	723	14	BM702190
5	69	2.0	496	17	AQ202422
6	59	1.7	567	10	AW855319

C	7	57	1.7	201	10	AW856235
	8	31	0.9	631	10	BB612001
	9	31	0.9	638	10	BB613951
	10	31	0.9	664	10	BB617396
	11	30	0.9	501	12	BF601265
	12	29	0.9	148	13	BG991167
C	13	25	0.7	395	9	AA910328
C	14	25	0.7	445	9	AI744019
C	15	25	0.7	474	9	AI793020
C	16	25	0.7	494	9	AI373743
C	17	25	0.7	557	10	AW241667
C	18	23	0.7	348	9	AI772957
C	19	23	0.7	519	9	AI018418
C	20	22	0.7	586	10	AW648529
C	21	22	0.7	204	10	BB571653
C	22	21	0.6	285	17	BH273274
C	23	21	0.6	300	9	AI991076
C	24	21	0.6	371	10	BE366992
C	25	21	0.6	371	17	AQ004063
C	26	21	0.6	469	9	AA103567
C	27	21	0.6	480	10	BE367090
C	28	21	0.6	490	9	AI668711
C	29	21	0.6	495	17	B50693
C	30	21	0.6	622	9	AI054920
C	31	21	0.6	632	17	A2344444
C	32	21	0.6	683	17	AQ957276
C	33	21	0.6	748	14	BM963047
C	34	21	0.6	750	14	BM947057
C	35	21	0.6	789	14	BQ443279
C	36	21	0.6	930	17	AQ749010
C	37	21	0.6	936	14	BQ941497
C	38	20	0.6	120	13	BM484625
C	39	20	0.6	194	9	AA078411
C	40	20	0.6	197	12	BF461151
C	41	20	0.6	197	12	BF461673
C	42	20	0.6	241	9	AA767572
C	43	20	0.6	241	17	AZ744203
C	44	20	0.6	256	10	AV361506
C	45	20	0.6	277	10	AW516927

ALIGNMENTS

RESULT 1
BF849816/c 535 bp mRNA linear EST 16-JAN-2001
LOCUS BF849816
DEFINITION PM4-EN0068-151100-004-b06 EN0068 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF849816
VERSION BF849816.1 GI:12236966
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 535)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&t2=PM4-EN0068-151100-004-b06&t3=2000-11-15&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 534.

FEATURES

Location/Qualifiers
1..535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="EN0068"
/dev_stage="Adult"
/note="Organ: lung_normal; Vector: puc18; Site: 1: SmaI; Site: 2: SmaI; A mini-library was made by cloning products derived from ORESFES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 142 a 150 c 124 g 119 t

Query Match . 9.0%; Score 305; DB 12; Length 535;
Best Local Similarity 100.0%; Pred. No. 3.7e-141; Indels 0; Gaps 0;
Matches 305; Conservative 0; Mismatches 0;
QY 2899 CCCACTTTCGTACCTGCTGTGCTGTTGAGGCTATGTCATCGCACCTTTCCTTGA 2958
Db 535 CCCACTTTCGTACCTGCTGTGCTGTTGAGGCTATGTCATCGCACCTTTCCTTGA 476
QY 2959 GGATAACAAGGGTCTGTAAGACTTAATTTAGCGGCTGAGCTTCCTTGCACACAAT 3018
Db 475 GGATAACAAGGGTCTGTAAGACTTAATTTAGCGGCTGAGCTTCCTTGCACACAAT 416
QY 3019 CAATGCTCGCCAGCAATGTTCTGCACAGTAATGCCAGCAGAGGCTTACTAGAGCAT 3078
Db 415 CAATGCTCGCCAGCAATGTTCTGCACAGTAATGCCAGCAGAGGCTTACTAGAGCAT 356
QY 3079 CTTTGGAGCGGCGAAGGCCAGCCCTTCAAGATGGAAGCAGCAGCTTTCACATTCCTCC 3138
Db 355 CTTTGGAGCGGCGAAGGCCAGCCCTTCAAGATGGAAGCAGCAGCTTTCACATTCCTCC 296
QY 3139 CAGACATCTCGATGTCATTTGCTGAGTCTGAAGGGGCTTGCAGGACCTTTGTGA 3198
Db 295 CAGACATCTCGATGTCATTTGCTGAGTCTGAAGGGGCTTGCAGGACCTTTGTGA 236
QY 3199 CTTCT 3203
Db 235 CTTCT 231

RESULT 2

AW206492/c 406 bp mRNA linear EST 02-DEC-1999
LOCUS UI-H-B11-aez-c-01-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens CDNA clone
DEFINITION IMAGE:2721001 3', mRNA sequence.

ACCESSION AW206492
VERSION AW206492.1 GI:6505988
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 406)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.

COMMENT

Email: cgabbs@rcmail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone Distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=yes.

FEATURES

Location/Qualifiers
1..406
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="EN0068"
/dev_stage="Adult"
/note="Organ: lung_normal; Vector: puc18; Site: 1: SmaI; Site: 2: SmaI; A mini-library was made by cloning products derived from ORESFES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 80 a 112 c 107 g 107 t

Query Match 8.8%; Score 296; DB 10; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.1e-136;
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2616 GATTCTTGAGACGCCCTGAGGCACATGGCTCCCGTCAGGAGGCGCAGCGGTACCCC 2675
Db 312 GATTCTTGAGACGCCCTGAGGCACATGGCTCCCGTCAGGAGGCGCAGCGGTACCCC 253
QY 2676 TCCCACTCACTACAGAGAGGCTGGGCACCTGAAATGGTGCCTACCTTCTGGAAATGCTG 2735
Db 252 TCCCACTCACTACAGAGAGGCTGGGCACCTGAAATGGTGCCTACCTTCTGGAAATGCTG 193
QY 2736 TGCCCCAGGTCCCTTAGAATGCTGCTTCCCGCGTGGCAGGACCACTATTCTCACTGAG 2795
Db 192 TGCCCCAGGTCCCTTAGAATGCTGCTTCCCGCGTGGCAGGACCACTATTCTCACTGAG 133
QY 2796 GGAGGAGGATGTCCCACTGCAGCCCATGCTGTAGAGACAAGAAAGCAGCTGATGTCAC 2855
Db 132 GGAGGAGGATGTCCCACTGCAGCCCATGCTGTAGAGACAAGAAAGCAGCTGATGTCAC 73

TAG_LIB=NCI_CGAP_Col0
TAG_TISSUE=colon
TAG_SEQ=AAACG"

BASE COUNT 80 a 112 c 107 g 107 t

ORIGIN

Qy 2856 CCACAACGATGTTGTGAAAGTTTGTGATGTGTAGTAATACCCACTTTCTCTA 2911
 |||||||
 Db 72 CCACAACGATGTTGTGAAAGTTTGTGATGTGTAGTAATACCCACTTTCTCTA 17
 |||||||

RESULT 3
 BE143292/c
 LOCUS
 DEFINITION MR0-HT0161-221099-002-c08 HT0161 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BE143292
 VERSION BE143292.1 GI:8606013
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 282)
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.J.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-MR0-HT0161-221
 099-002-c08&t3=1999-10-22&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 8
 High quality sequence stop: 282.

FEATURES
 source
 1..282
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HT0161"
 /dev_stage="Adult"
 /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 71 a 78 c 63 g 70 t
 ORIGIN
 Query Match 3.8%; Score 128; DB 10; Length 282;
 Best Local Similarity 100.0%; Pred. No. 1.7e-52;
 Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 3026 CGCCAGATGTTGTGACAGTAATGCCAGCAGAGCCCTTTACTAGAGCATCCTTTGG 3085
 |||||||
 Db 282 CGCCAGATGTTGTGACAGTAATGCCAGCAGAGCCCTTTACTAGAGCATCCTTTGG 223
 |||||||
 Qy 3086 ACGCGAAGCCAGCGCCTTTCAAGATGGAAGCAGCAGCTTTTCCACTTCCCCAGAGAC 3145
 |||||||
 Db 222 ACGCGAAGCCAGCGCCTTTCAAGATGGAAGCAGCAGCTTTTCCACTTCCCCAGAGAC 163
 |||||||
 Qy 3146 ATTCTGGA 3153
 |||||||
 Db 162 ATTCTGGA 155
 |||||||

RESULT 4
 BM702190/c
 LOCUS
 DEFINITION UI-E-CQ1-aey-m-08-0-UI.r1 UI-E-CQ1 Homo sapiens cDNA clone
 UI-E-CQ1-aey-m-08-0-UI 5', mRNA sequence.
 ACCESSION BM702190
 VERSION BM702190.1 GI:19015448
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 723)
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 Cloning performed by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 The following repetitive elements were found in this cDNA
 sequence: 353-608, >LINE2
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..723
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-CQ1-aey-m-08-0-UI"
 /clone_lib="UI-E-CQ1"
 /tissue_type="optic nerve"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT7T3-pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-CQ1 is a normalized cDNA library containing the
 following tissue(s): optic nerve. The library was
 constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into pT7T3-pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 CCATTAAGTC. This library was created for the program, Gene
 Discovery in the Visual System, supported by National Eye
 Institute (NEI)."
 BASE COUNT 147 a 241 c 142 g 190 t 3 others
 ORIGIN
 Query Match 3.3%; Score 113; DB 14; Length 723;
 Best Local Similarity 100.0%; Pred. No. 6.6e-45;
 Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2615 GGATTTCTGAGAGCCCTCGAGGCATGCTCCCGTCGAGAGGGGCGAGCGGTACCC 2674
 |||||||
 Db 113 GGATTTCTGAGAGCCCTCGAGGCATGCTCCCGTCGAGAGGGGCGAGCGGTACCC 54
 |||||||
 Qy 2675 CTCCCAGCAACTACAGAGAGGCGCTGGGCGACTGAAATGGTGTACCTTCTGG 2727
 |||||||


```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="5031412M06"
/clone_lib="RIKEN full-length enriched, 11 days pregnant
adult female ovary and uterus"
/sex="female"
/tissue_type="ovary and uterus"
/dev_stage="11 days pregnant, adult"
/lab_host="DH10B"
/notes="Site_1: Sali; Site_2: BamHI. cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGGATCCAGAGCTCTTTTTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATCTCGAGTTAATTAATATCCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end:
BamHI."
BASE COUNT 134 a 196 c 195 g 139 t
ORIGIN

Query Match 0.9%; Score 31; DB 10; Length 664;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 614 GTCCACTCTCATCTCGGATTCCTCGA 644
|||||
Db 363 GTCCACTCTCATCTCGGATTCCTCGA 393

RESULT 11
BF601265 501 bp mRNA linear EST 25-APR-2001
LOCUS
DEFINITION 266197 MARC 380V Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF601265
VERSION BF601265.1 GI:11698487
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 501)
Cass, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett
, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G.,
Perteau, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keefe, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR PRIMERS
FORWARD: AGGAACAGCTATGACCAT

```

```

BACKWARD: GTTTCCTCCAGTCACGACG
Plate: 40 row: J column: 10
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..501
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 380V"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: pCMV SPORT6; Site_1: NotI; Site_2: Sali;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT 98 a 164 c 154 g 85 t
ORIGIN

Query Match 0.9%; Score 30; DB 12; Length 501;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2547 TGGCTGGAGGGCCCCCTGCGAGAACG 2576
|||||
Db 24 TGGCTGGAGGGCCCCCTGCGAGAACG 53

RESULT 12
BG991167 148 bp mRNA linear EST 13-JUN-2001
LOCUS
DEFINITION MR2-HT1189-270101-003-h07 HT1189 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG991167
VERSION BG991167.1 GI:14395237
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 148)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2st2-MR2-HT1189-
270101-003-h07st3-2001-01-27st4-1)
Seq primer: puc 18 forward
High quality sequence start: 33
High quality sequence stop: 147.
Location/Qualifiers
1..148
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT1189"
/dev_stage="Adult"
/notes="Organ: head neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)

```

profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 29 a 43 c 47 g 29 t

ORIGIN

Query Match 0.7%; Score 29; DB 13; Length 148;
Best Local Similarity 100.0%; Pred. No. 0.0064;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2743 GTCCTTAGAATCTGCTCCGCGGTG 2771

Db 42 GTCCTTAGAATCTGCTCCGCGGTG 70

RESULT 13

AA910328/c

LOCUS AA910328 395 bp mRNA linear EST 13-APR-1998
DEFINITION ok83c11.s1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1520564 3'
similar to TR:Q23242 Q23242 ZC116.3 i, mRNA sequence.

ACCESSION AA910328

VERSION AA910328.1 GI:3049618

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 395)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40m13 fwd EF from Amersham.

FEATURES

source

1. 395

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1520564"

/clone_lib="NCI_CGAP_Kid3"

/lab_host="DH10B"

/note="organ: kidney; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer,

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not

I and Eco RI sites of the modified pT73 vector. mRNA

source: 2 pooled kidneys. Library went through one round

of normalization. Library constructed by Bento Soares and

M. Fatima Bonaldo."

BASE COUNT 90 a 101 c 84 g 120 t

ORIGIN

Query Match

Best Local Similarity 0.7%; Score 25; DB 9; Length 395;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1236 CTGCCAGATGGAGGCACATGTGTT 1260

Db 333 CTGCCAGATGGAGGCACATGTGTT 309

RESULT 14

AI744019/c

LOCUS

DEFINITION

AI744019 wc32g02.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2316914 3'

similar to TR:O60494 O60494 INTRINSIC FACTOR-B12 RECEPTOR

PRECURSORS: i, mRNA sequence.

ACCESSION AI744019

VERSION AI744019.1 GI:5112307

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 445)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 589 Std Error: 0.00

Seq primer: -40UP from Gibco.

FEATURES

source

1. 445

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2316914"

/clone_lib="NCI_CGAP_Kid11"

/lab_host="DH10B"

/note="organ: kidney; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

Plasmid DNA from the normalized library NCI_CGAP_Kid3 was

prepared, and ss circles were made in vitro. Following HAP

hybridization, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5 000 clones made from the same library

(cloneids 1322376-1323911, 1456007-1456775, and

1500552-1502855). Subtraction by Bento Soares and M.

Fatima Bonaldo."

BASE COUNT 107 a 111 c 95 g 132 t

ORIGIN

Query Match 0.7%; Score 25; DB 9; Length 445;

Best Local Similarity 100.0%; Pred. No. 0.77;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1236 CTGCCAGATGGAGGCACATGTGTT 1260

Db 344 CTGCCAGATGGAGGCACATGTGTT 320

RESULT 15

AI793020

LOCUS

DEFINITION

AI793020 q239a03.y5 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2029228 5'

similar to TR:O60494 O60494 INTRINSIC FACTOR-B12 RECEPTOR

PRECURSORS: i, mRNA sequence.

ACCESSION AI793020

VERSION AI793020.1 GI:5340736

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 474)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index
Unpublished (1997)
Other ESTs: qz39a03.xl
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-re@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbr/image/image.ntml

This read is a RESSEQUENCE of a previously sequenced human clone
Original clone citation: see original entry for original citation
information

This 5' resequenced clone has no previous 5' data to verify this
new read against
Insert Length: 878 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 425.

FEATURES

Location/Qualifiers

Source

1..474
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2029228"
/clone_lib="NCI_CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT 148 a 89 c 107 g 130 t

ORIGIN

Query Match 0.78; Score 25; DB 9; Length 474;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1236 CTGCCAGATGGAGCACATGTGT 1260

Db 411 CTGCCAGATGGAGCACATGTGT 435

Search completed: May 5, 2003, 07:32:42
Job time : 3012 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 22:50:07 ; Search time 5749 Seconds
(without alignments)
17085.053 Million cell updates/sec

Title: US-09-930-020A-1

Perfect score: 3375

Sequence: 1 gacagtgttcgagctgcac.....tgtccacctgaaggtcttc 3375

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_mn.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pi.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_mn.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pi.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	712.6	21.1	123110	9	AC005383	AC005383 Homo sapi
2	408.6	12.1	152895	2	AC119647	AC119647 Rattus no
3	408.6	12.1	154748	2	AC125150	AC125150 Mus muscu
C 4	317.8	9.4	160420	9	AC022023	AC022023 Homo sapi
C 5	309.8	9.2	136357	2	AL161942	AL161942 Homo sapi
C 6	123.4	3.7	65824	2	AC099899	AC099899 Mus muscu
C 7	111.8	3.3	65824	2	AC099899	AC099899 Mus muscu
C 8	97.8	2.9	4151	6	AX281619	AX281619 Sequence
C 9	96	2.8	152895	2	AC119647	AC119647 Rattus no
C 10	95.6	2.8	1304	5	CHKMP	M14792 Chicken car
C 11	95.6	2.8	3327	9	HUMCMPMR	N55683 Human carti
C 12	87.8	2.6	7504	9	AB040943	AB040943 Homo sapi
C 13	86.8	2.6	194143	2	AC112730	AC112730 Rattus no
C 14	85	2.5	1959	10	MMU35035	U35035 Mus musculu
C 15	82	2.4	190669	10	AL590429	AL590429 Mouse DNA
C 16	80	2.4	3571	10	MMU69262	U89262 Mus musculu
C 17	79.4	2.4	9235	10	MMU32107	U32107 Mus musculu
C 18	78.6	2.3	2193	10	MMU6140	AJ006140 Mus muscu
C 19	78.4	2.3	3273	10	BC005429	BC005429 Mus muscu
C 20	77	2.3	169585	2	AC078821	AC078821 Homo sapi
C 21	76.6	2.3	186666	2	AC097136	AC097136 Rattus no
C 22	76.4	2.3	177695	2	AC096154	AC096154 Rattus no
C 23	76.4	2.3	192817	2	AC127063	AC127063 Rattus no
C 24	75.4	2.2	1360	9	HUMCMP5	M55679 Human carti
C 25	75.4	2.2	152042	9	AL137857	AL137857 Human DNA
C 26	75.2	2.2	3550	9	BC010444	BC010444 Homo sapi
C 27	74.8	2.2	202006	2	AL669980	AL669980 Mus muscu
C 28	74.2	2.2	1167	9	AK027323	AK027323 Homo sapi
C 29	74.2	2.2	2003	6	HSA17581	AJ007581 Homo sapi
C 30	73.6	2.2	1033	6	AX329755	AX329755 Sequence
C 31	73.6	2.2	2293	9	AK027775	AK027775 Homo sapi
C 32	73.6	2.2	2733	9	BC016394	BC016394 Homo sapi
C 33	73.6	2.2	2745	9	HSM802406	AL137638 Homo sapi
C 34	73.6	2.2	2863	6	AX015418	AX015418 Sequence
C 35	73.6	2.2	3032	6	AX399978	AX399978 Sequence
C 36	73.6	2.2	3089	6	AX399977	AX399977 Sequence
C 37	73.6	2.2	3373	6	AR085071	AR085071 Sequence
C 38	73.6	2.2	3476	6	AX079878	AX079878 Sequence
C 39	73.6	2.2	3449	6	AX464160	AX464160 Sequence
C 40	73.6	2.2	3496	9	HSU69263	U69263 Homo sapien
C 41	73.6	2.2	3572	9	HSM805245	AL833931 Homo sapi
C 42	73.6	2.2	147620	9	HS453C12	AL021578 Human DNA
C 43	73.6	2.2	189656	9	AF314058	AF314058 Homo sapi
C 44	73.4	2.2	595	5	GGCMP5	X12350 Chicken car
C 45	72.6	2.2	3026	6	AX079876	AX079876 Sequence

ALIGNMENTS

RESULT 1
AC005383
LOCUS AC005383 Homo sapiens chromosome 10 clone C1987SK-114466 map 10q25.1,
DEFINITION complete sequence.
ACCESSION AC005383
VERSION AC005383.1 GI:3818355
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 123110)
AUTHORS Smith,D.R.
TITLE Sequencing of Human Chromosome 10

AC005383 AC005383 123110 bp DNA linear PRI 31-OCT-1998
Homo sapiens chromosome 10 clone C1987SK-114466 map 10q25.1,
complete sequence.

```
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 123110)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (05-AUG-1998) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02154, USA
REFERENCE 3 (bases 1 to 123110)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-1998) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02154, USA
COMMENT On Oct 31, 1998 this sequence version replaced gi:3808081.
FEATURES
    source
        location/Qualifiers
            1..123110
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="10"
                /map="10q25.1"
                /clone="C1987SK-1144G6"
BASE COUNT 31394 a 28668 c 29634 g 33413 t 1 others
ORIGIN
Query Match 21.1%; Score 712.6; DB 9; Length 123110;
Best Local Similarity 98.4%; Pred. No. 4.2e-141;
Matches 751; Conservative 0; Mismatches 9; Indels 3; Gaps 3;
QY 2615 GGATTCCTTGAGAGCCCTCGAGGACATGCTCCCGTCGAGGAGGCGAGCGGTACCC 2674
Db 78153 GGATTCCTTGAGAGCCCTCGAGGACATGCTCCCGTCGAGGAGGCGAGCGGTACCC 78212
QY 2675 CTCGCCAGCACTACAGAGAGCGCTGGCACTGAATGTGGCTACTCTTCGTAATGCTCT 2734
Db 78213 CTCGCCAGCACTACAGAGAGCGCTGGCACTGAATGTGGCTACTCTTCGTAATGCTCT 78272
QY 2735 GTGCCCGAGCTCTTAGAATGTCTGCTTCCCGCGCTGGCCGAGGACCACTATTCTCACTGA 2794
Db 78273 GTGCCCGAGCTCTTAGAATGTCTGCTTCCCGCGCTGGCCGAGGACCACTATTCTCACTGA 78332
QY 2795 GGAGGAGGATGCCCACTGACGACATGCTGCTGTAGAGACAGAAAGCAGCTGATGCTCA 2854
Db 78333 GGAGGAGGATGCCCACTGACGACATGCTGCTGTAGAGACAGAAAGCAGCTGATGCTCA 78392
QY 2855 CCCACAACCATGTTGTTGAAAAGTTTGTATGTGTAAGTAATACCCACTTTCTGTACCT 2914
Db 78393 CCCACAACCATGTTGTTGAAAAGTTTGTATGTGTAAGTAATACCCACTTTCTGTACCT 78452
QY 2915 GCTGTGCTTTGTGAGGCTATGTCATCTGCCACCTTTCCCTTGAGGATAAACAAGGGGTC 2974
Db 78453 GCTGTGCTTTGTGAGGCTATGTCATCTGCCACCTTTCCCTTGAGGATAAACAAGGGGTC 78512
QY 2975 CTGAAGACTTAAATTTAGCGGCTGAGCTTCTTTTGACACAATCAATGTCGCCAGAT 3034
Db 78513 CTGAAGACTTAAATTTAGCGGCTGAGCTTCTTTTGACACAATCAATGTCGCCAGAT 78572
QY 3035 GTTCTTCACACAGTAATGCCACAGAGCGCTTTACTAGAGCATCTTTTGAGCGGCGAAG 3094
Db 78573 GTTCTTCACACAGTAATGCCACAGAGCGCTTTACTAGAGCATCTTTTGAGCGGCGAAG 78632
QY 3095 GCCACGCGCTTTCAAGATGGAAGACAGCAGCTTTTCCACATCTCCCGAGACATCTTGGAT 3154
Db 78633 GCCACGCGCTTTCAAGATGGAAGACAGCAGCTTTTCCACATCTCCCGAGACATCTTGGAT 78692
QY 3155 GCATTTGCATGTAGCTGAAGGGGGCTTGAGGAGCTTTGTGACTCTTTGGGCACTGCC 3214
Db 78693 GCATTTGCATGTAGCTGAAGGGGGCTTGAGGAGCTTTGTGACTCTTTGGGCACTGCC 78751
QY 3215 TTTTGTGTGAGAGACACTTGAAGGCTTCACACTG-AATGTGACCAATTAAACAGCT 3273
Db 78752 TTTTGTGTGAGAGACACTTGAAGGCTTCACACTG-AATGTGACCAATTAAACAGCT 78811
QY 3274 TGGTTGATGATGGGGGAGGGGCTGAGTTGTGCATGGCCCGCCAGGTCTGGA-GGGCCACGTA 3332
Db 78812 TGGTTGATGATGGGGGAGGGGCTGAGTTGTGCATGGCCCGCCAGGTCTGGA-GGGCCACGTA 78871
```

```
QY 3333 AAATCCTTCTGAGTCGTGAGCAGTGTCCACCTTGAAGGCTTCC 3375
|||||
Db 78872 AAATCCTTCTGAGTCGTGAGCAGTGTCCACCTTGAAGGCTTCC 78914
```

RESULT 2

```
AC119647
LOCUS
```

```
DEFINITION
```

```
***, 54 unordered pieces.
```

```
AC119647.2 GI:21747172
```

```
HTG: HTGS-PHASE1.
```

```
Rattus norvegicus
```

```
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
```

```
Rattus.
```

```
1 (bases 1 to 152895)
```

```
Muzdy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
```

```
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
```

```
Barbarta,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
```

```
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
```

```
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
```

```
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
```

```
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
```

```
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
```

```
Dunaway,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
```

```
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
```

```
Falls,T., Ferraguto,D., Flagg,N., Foster,J., Foster,P., Frantz,P.,
```

```
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
```

```
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
```

```
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
```

```
Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
```

```
Jacobson,B., Jia,X., Johnson,R., Jolivet,S., Joudan,S.,
```

```
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
```

```
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
```

```
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseg,H.,
```

```
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
```

```
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
```

```
Massey,B., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
```

```
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
```

```
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
```

```
Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okwona,G.,
```

```
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
```

```
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
```

```
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
```

```
Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sissob,I.,
```

```
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
```

```
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
```

```
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
```

```
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
```

```
Williams,G., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
```

```
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
```

```
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
```

```
Weinstock,G. and Gibbs,R.
```

```
Direct Submission
```

```
Unpublished
```

```
2 (bases 1 to 152895)
```

```
Worley,K.C.
```

```
Direct Submission
```

```
Submitted (30-APR-2002) Human Genome Sequencing Center, Department
```

```
of Molecular and Human Genetics, Baylor College of Medicine, One
```

```
Baylor Plaza, Houston, TX 77030, USA
```

```
3 (bases 1 to 152895)
```

```
Worley,K.C.
```

```
Direct Submission
```

```
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
```

```
of Molecular and Human Genetics, Baylor College of Medicine, One
```


COMMENT
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20340380.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVPZ
Center clone name: CH230-137H14
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 91714 bases at least Q40
Consensus quality: 98583 bases at least Q30
Consensus quality: 103878 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 54 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1
1042: 1041: contig of 1041 bp in length
1141: 1141: gap of unknown length
1142: 2262: contig of 1121 bp in length
2263: 2362: gap of unknown length
2363: 3497: contig of 1135 bp in length
3498: 3597: gap of unknown length
3598: 5210: contig of 1613 bp in length
5211: 5310: gap of unknown length
5311: 6936: contig of 1626 bp in length
6937: 7036: gap of unknown length
7037: 8291: contig of 1255 bp in length
8292: 8963: gap of unknown length
8964: 9763: gap of unknown length
9764: 11369: contig of 1606 bp in length
11370: 11469: gap of unknown length
11470: 13073: contig of 1604 bp in length
13074: 13173: gap of unknown length
13174: 14521: contig of 1348 bp in length
14522: 14621: gap of unknown length
14622: 16274: contig of 1653 bp in length
16275: 16374: gap of unknown length
16376: 17405: contig of 1031 bp in length
17406: 17505: gap of unknown length
17506: 18547: contig of 1042 bp in length
18548: 18647: gap of unknown length
18649: 19908: contig of 1261 bp in length
19909: 20008: gap of unknown length
20009: 21460: contig of 1452 bp in length
21461: 21560: gap of unknown length
21561: 22822: contig of 1262 bp in length
22823: 22922: gap of unknown length
22923: 24437: contig of 1515 bp in length
24438: 24537: gap of unknown length
24539: 25832: contig of 1295 bp in length
25833: 25932: gap of unknown length
25934: 27104: contig of 1172 bp in length
27105: 27204: gap of unknown length
27206: 29052: contig of 1848 bp in length
29053: 29152: gap of unknown length
29153: 30398: contig of 1246 bp in length
30399: 30498: gap of unknown length
30499: 32426: contig of 1928 bp in length
32427: 32526: gap of unknown length
32527: 34955: contig of 2429 bp in length
```

```
34956: 35055: gap of unknown length
35056: 36642: contig of 1587 bp in length
36643: 36643: gap of unknown length
36743: 37820: contig of 1078 bp in length
37821: 37920: gap of unknown length
37921: 39394: contig of 1474 bp in length
39395: 39494: gap of unknown length
39495: 40928: contig of 1434 bp in length
40929: 41028: gap of unknown length
41029: 43380: contig of 2352 bp in length
43381: 43480: gap of unknown length
43481: 44748: contig of 1268 bp in length
44749: 44848: gap of unknown length
44849: 46927: contig of 2079 bp in length
46928: 47027: gap of unknown length
47029: 49481: contig of 2454 bp in length
49482: 49581: gap of unknown length
49582: 51923: contig of 2342 bp in length
51924: 52023: gap of unknown length
52024: 54126: contig of 2103 bp in length
54127: 54226: gap of unknown length
54227: 57214: contig of 2988 bp in length
57215: 57314: gap of unknown length
57316: 60024: contig of 2710 bp in length
60025: 60124: gap of unknown length
60126: 63163: contig of 3039 bp in length
63164: 63263: gap of unknown length
63264: 66703: contig of 3440 bp in length
66704: 66803: gap of unknown length
66804: 68557: contig of 1754 bp in length
68558: 68657: gap of unknown length
68659: 72882: contig of 4225 bp in length
72883: 72982: gap of unknown length
72984: 76542: contig of 3560 bp in length
76543: 76642: gap of unknown length
76644: 80420: contig of 3778 bp in length
80421: 80520: gap of unknown length
80521: 84990: contig of 4470 bp in length
84991: 85090: gap of unknown length
85091: 88933: contig of 3843 bp in length
88934: 89033: gap of unknown length
89034: 96304: contig of 7271 bp in length
96305: 96404: gap of unknown length
96406: 101068: contig of 4664 bp in length
101069: 101168: gap of unknown length
101169: 105041: contig of 3873 bp in length
105042: 105141: gap of unknown length
105142: 110814: contig of 5673 bp in length
110815: 110914: gap of unknown length
110915: 115458: contig of 4544 bp in length
115459: 115558: gap of unknown length
115559: 120911: contig of 5353 bp in length
120912: 121011: gap of unknown length
121012: 128307: contig of 7296 bp in length
128308: 128407: gap of unknown length
128408: 134219: contig of 5812 bp in length
134220: 134319: gap of unknown length
134320: 139473: contig of 5154 bp in length
139474: 139573: gap of unknown length
139574: 146170: contig of 6597 bp in length
146171: 146270: gap of unknown length
146271: 152895: contig of 6625 bp in length.
```

```
Query Match      12.1%  Score 408.6;  DB 2;  Length 152895;
Best Local Similarity 81.9%;  Pred. No. 1.9e-76;
Matches 471;  Conservative 0;  Mismatches 104;  Indels 0;  Gaps 0;

Qy 1324  GCCCTGAAGCTGAGCCCTGGGAATGAGGCTGACCTCTCTCTCTCTGCGACAGCTCTGCG 1383
      |||||
Db 85817  GCCCGAAGCTGAGCCCTGGGAATGAGGCTGACCTCTCTCTCTCTGCGACAGCTCTGCA 85876
      |||||

Qy 1384  GGCACCACTCTGGACGCTTCTCGGGCCCAAGCTTCGGAAGCGTTTGTGCGGCC 1443
      |||||
```


* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 5425: contig of 5425 bp in length
5426 5525: gap of 100 bp
5526 9483: contig of 3958 bp in length
9484 9583: gap of 100 bp
9584 14280: contig of 4697 bp in length
14281 14380: gap of 100 bp
14381 17312: contig of 2932 bp in length
17313 17412: gap of 100 bp
17413 22848: contig of 5436 bp in length
22849 22948: gap of 100 bp
22949 26090: contig of 3142 bp in length
26091 26190: gap of 100 bp
26191 28215: contig of 2025 bp in length
28216 28315: gap of 100 bp
28316 33862: contig of 5547 bp in length
33863 33962: gap of 100 bp
33963 44145: contig of 10183 bp in length
44146 44245: gap of 100 bp
44246 53165: contig of 8920 bp in length
53166 53265: gap of 100 bp
53266 57412: contig of 4147 bp in length
57413 57512: gap of 100 bp
57513 60170: contig of 2658 bp in length
60171 60270: gap of 100 bp
60271 65985: contig of 5715 bp in length
65986 66085: gap of 100 bp
66086 69094: contig of 3009 bp in length
69095 69194: gap of 100 bp
69195 71736: contig of 2542 bp in length
71737 71836: gap of 100 bp
71837 74955: contig of 3119 bp in length
74956 75055: gap of 100 bp
75056 77325: contig of 2270 bp in length
77326 77425: gap of 100 bp
77426 80709: contig of 3284 bp in length
80710 80809: gap of 100 bp
80810 82943: contig of 2134 bp in length
82944 83043: gap of 100 bp
83044 85218: contig of 2175 bp in length
85219 85318: gap of 100 bp
85319 89861: contig of 4543 bp in length
89862 89961: gap of 100 bp
89962 92824: contig of 2863 bp in length
92825 92924: gap of 100 bp
92925 95822: contig of 2898 bp in length
95823 95922: gap of 100 bp
95923 98130: contig of 2208 bp in length
98131 98230: gap of 100 bp
98231 100594: contig of 2364 bp in length
100595 100694: gap of 100 bp
100695 109043: contig of 8349 bp in length
109044 109143: gap of 100 bp
109144 112146: contig of 3003 bp in length
112147 112246: gap of 100 bp
112247 117183: contig of 4937 bp in length
117184 117283: gap of 100 bp
117284 121861: contig of 4578 bp in length
121862 121961: gap of 100 bp
121962 124258: contig of 2297 bp in length
124259 124358: gap of 100 bp
124359 127036: contig of 2678 bp in length
127037 127136: gap of 100 bp
127137 133357: contig of 6221 bp in length
133358 133457: gap of 100 bp
133458 136357: contig of 2900 bp in length.

FEATURES
Source

Location/Qualifiers
1. .136357
/organism="Homo sapiens"
/db_xref="taxon:9606"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

/chromosome="20"
/clone="RP11-83B5"
/clone_lib="RPC1-11.1"
1. 5425
/note="assembly_fragment:01370"
fragment_chain:1
clone_end:SP6
vector_side:left
5526..9483
/note="assembly_fragment:01120"
fragment_chain:1
9584..14280
/note="assembly_fragment:00095"
fragment_chain:1
14381..17312
/note="assembly_fragment:01155"
fragment_chain:1
17413..22848
/note="assembly_fragment:00025"
fragment_chain:2
22949..26090
/note="assembly_fragment:00387"
fragment_chain:2
26191..28215
/note="assembly_fragment:00262"
fragment_chain:3
28316..33862
/note="assembly_fragment:00134"
fragment_chain:3
33963..44145
/note="assembly_fragment:00605"
fragment_chain:4
44246..53165
/note="assembly_fragment:00989"
fragment_chain:4
53266..57412
/note="assembly_fragment:00876"
fragment_chain:5
57513..60170
/note="assembly_fragment:00248"
fragment_chain:5
60271..65985
/note="assembly_fragment:01554"
fragment_chain:6
66086..69094
/note="assembly_fragment:00470"
fragment_chain:6
69195..71736
/note="assembly_fragment:01566"
fragment_chain:7
71837..74955
/note="assembly_fragment:00477"
fragment_chain:7
75056..77325
/note="assembly_fragment:00036"
77426..80709
/note="assembly_fragment:00390"
80810..82943
/note="assembly_fragment:00409"
83044..85218
/note="assembly_fragment:00437"
85319..89861
/note="assembly_fragment:00764"
89962..92824
/note="assembly_fragment:00782"
92925..95822
/note="assembly_fragment:00808"
95923..98130
/note="assembly_fragment:00831"
98231..100594
/note="assembly_fragment:00886"
100695..109043
/note="assembly_fragment:00992"

```

misc_feature 109144..112146
              /note="assembly_fragment:01007"
misc_feature 112247..117183
              /note="assembly_fragment:01102"
misc_feature 117284..121861
              /note="assembly_fragment:01146"
misc_feature 121962..124258
              /note="assembly_fragment:01320"
misc_feature 124359..127036
              /note="assembly_fragment:01489"
misc_feature 127137..133357
              /note="assembly_fragment:01498"
misc_feature 133458..136357
              /note="assembly_fragment:01604"
BASE COUNT 38044 a 27571 c 27577 g 39953 t 3212 others
ORIGIN

```

```

Query Match          9.28; Score 309.8; DB 2; Length 136357;
Best Local Similarity 93.68; Pred. No. 1.9e-55;
Matches 323; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 2 ACAGTGTTCGGGGTGCACCGCTCGGAGCTGGGTGAGCCGCGTAGAAGTGAAGTACTTT 61
    |||||||
Db 9069 ACAGTGTTCGGGGTGCACCGCTCGGAGCTGGGTGAGCCGCGTAGAAGTGAAGTACTTT 9010
    |||||||

Qy 62 TTTATTTTCAGACCTGGCGGCGATGCCGCTTTTAAAAAAGCGAGGGGCTCTATGCACCTCC 121
    |||||||
Db 9009 TTTATTTTCAGACCTGGCGGCGATGCCGCTTTTAAAAAAGCGAGGGGCTCTATGCACCTCC 8950
    |||||||

Qy 122 CTGGCGGTAGTTCCTCCACCTCAGCCGGGTGCGGTGCGGCCCTCTCCAGAGAGA 181
    |||||||
Db 8949 CTGGCGGTAGTTCCTCCACCTCAGCCGGGTGCGGTGCGGCCCTCTCCAGAGAGA 8890
    |||||||

Qy 182 CAACACAGTGTCCACGTGGCAGCGCCGCCGCCCTCTCTGTGATCCCGTAGCGC 241
    |||||||
Db 8889 CAACACAGTGTCCACGTGGCAGCGCCGCCGCCCTCTCTGTGATCCCGTAGCGC 8830
    |||||||

Qy 242 CCCCTGGCCCGAGCGCGCCCGGGTCTGTGAGTAGAGCCGCCCGGAGCGGCGCTG 301
    |||||||
Db 8829 CCCCTGGCCCGAGCGCGCCCGGGTCTGTGAGTAGAGCCGCCCGGAGCGGCGCTG 8770
    |||||||

Qy 302 CGCCGCTCTCTCTCGTTATACATGCCCTCTCTCTCTG 346
    |||||
Db 8769 CGCCGCTCTCTCTCGTTATACATGCCCTCTCTCTG 8725
    |||||

```

```

RESULT 6
AC099899 AC099899 65824 bp DNA linear HTG 22-NOV-2001
LOCUS Mus musculus clone RP23-12A20, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC099899
ACCESSION AC099899
VERSION AC099899.1 GI:17047265
KEYWORDS HTG; HTGS. PHASED.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 65824)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-12A20
Unpublished
REFERENCE 2 (bases 1 to 65824)
  Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
  Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
  Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
  Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
  Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
  Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
  Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
  Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
  Jones,C., Kamat,A., Karatas,A., Kells,C., Labrecque,K.,
  Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
  MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,

```

```

McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L13492
Center clone name: 12_A_20
-----
* NOTE: This record contains 82 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 689: contig of 689 bp in length
* 690 789: gap of 100 bp
* 790 1505: contig of 716 bp in length
* 1506 1605: gap of 100 bp
* 1606 2298: contig of 693 bp in length
* 2299 2398: gap of 100 bp
* 3093 3192: contig of 694 bp in length
* 3193 3864: contig of 672 bp in length
* 3865 3964: gap of 100 bp
* 3965 4677: contig of 713 bp in length
* 4678 4777: gap of 100 bp
* 4778 5469: contig of 692 bp in length
* 5470 5569: gap of 100 bp
* 5570 6268: contig of 699 bp in length
* 6269 6368: gap of 100 bp
* 6369 7088: contig of 720 bp in length
* 7089 7188: gap of 100 bp
* 7189 7901: contig of 713 bp in length
* 7902 8001: gap of 100 bp
* 8002 8705: contig of 704 bp in length
* 8706 8805: gap of 100 bp
* 8806 9503: contig of 698 bp in length
* 9504 9603: gap of 100 bp
* 9604 10289: contig of 686 bp in length
* 10290 10389: gap of 100 bp
* 10390 11101: contig of 712 bp in length
* 11102 11201: gap of 100 bp
* 11202 11908: contig of 707 bp in length
* 11909 12008: gap of 100 bp
* 12009 12714: contig of 706 bp in length
* 12715 12814: gap of 100 bp
* 12815 13516: contig of 702 bp in length
* 13517 13616: gap of 100 bp
* 13617 14321: contig of 705 bp in length
* 14322 14421: gap of 100 bp

```

TITLE
JOURNAL
COMMENT

Ginde,S., Gord,S., Goyette,M., Graham,L., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mhova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L13492
Center clone name: 12_A_20

* NOTE: This record contains 82 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
* 690 789: gap of 100 bp
* 790 1505: contig of 716 bp in length
* 1506 1605: gap of 100 bp
* 1606 2298: contig of 693 bp in length
* 2299 2398: gap of 100 bp
* 2399 3092: contig of 694 bp in length
* 3093 3192: gap of 100 bp
* 3193 3864: contig of 672 bp in length
* 3865 3964: gap of 100 bp
* 3965 4677: contig of 713 bp in length
* 4678 4777: gap of 100 bp
* 4778 5469: contig of 692 bp in length
* 5470 5569: gap of 100 bp
* 5570 6268: contig of 699 bp in length
* 6269 6368: gap of 100 bp
* 6369 7088: contig of 720 bp in length
* 7089 7188: gap of 100 bp
* 7189 7901: contig of 713 bp in length
* 7902 8001: gap of 100 bp
* 8002 8705: contig of 704 bp in length
* 8706 8805: gap of 100 bp
* 8806 9503: contig of 698 bp in length
* 9504 9603: gap of 100 bp
* 9604 10289: contig of 686 bp in length
* 10290 10389: gap of 100 bp
* 10390 11101: contig of 712 bp in length
* 11102 11201: gap of 100 bp
* 11202 11908: contig of 707 bp in length
* 11909 12008: gap of 100 bp
* 12009 12714: contig of 706 bp in length

* 12715 12814: gap of 100 bp
* 12815 13516: contig of 702 bp in length
* 13517 13616: gap of 100 bp
* 13617 14321: contig of 705 bp in length
* 14322 14421: gap of 100 bp
* 14422 15117: contig of 696 bp in length
* 15118 15217: gap of 100 bp
* 15218 15922: contig of 705 bp in length
* 15923 16022: gap of 100 bp
* 16023 16734: contig of 712 bp in length
* 16735 16834: gap of 100 bp
* 16835 17543: contig of 709 bp in length
* 17544 17643: gap of 100 bp
* 17644 18339: contig of 696 bp in length
* 18340 18439: gap of 100 bp
* 18440 19142: contig of 703 bp in length
* 19143 19242: gap of 100 bp
* 19243 19956: contig of 714 bp in length
* 19957 20056: gap of 100 bp
* 20057 20780: contig of 724 bp in length
* 20781 20880: gap of 100 bp
* 20881 21599: contig of 719 bp in length
* 21600 21699: gap of 100 bp
* 21700 22381: contig of 682 bp in length
* 22382 22481: gap of 100 bp
* 22482 23192: contig of 711 bp in length
* 23193 23292: gap of 100 bp
* 23293 23998: contig of 706 bp in length
* 23999 24098: gap of 100 bp
* 24099 24807: contig of 709 bp in length
* 24808 24907: gap of 100 bp
* 24908 25602: contig of 695 bp in length
* 25603 25702: gap of 100 bp
* 25703 26316: contig of 614 bp in length
* 26317 26416: gap of 100 bp
* 26417 27125: contig of 709 bp in length
* 27126 27225: gap of 100 bp
* 27226 27934: contig of 709 bp in length
* 27935 28034: gap of 100 bp
* 28035 28757: contig of 723 bp in length
* 28758 28857: gap of 100 bp
* 28858 29554: contig of 697 bp in length
* 29555 29654: gap of 100 bp
* 29655 30358: contig of 704 bp in length
* 30359 30458: gap of 100 bp
* 30459 31165: contig of 707 bp in length
* 31166 31265: gap of 100 bp
* 31266 31970: contig of 705 bp in length
* 31971 32070: gap of 100 bp
* 32071 32770: contig of 700 bp in length
* 32771 32870: gap of 100 bp
* 32871 33578: contig of 708 bp in length
* 33579 33678: gap of 100 bp
* 33679 34390: contig of 712 bp in length
* 34391 34490: gap of 100 bp
* 34491 35199: contig of 709 bp in length
* 35200 35299: gap of 100 bp
* 35300 35998: contig of 699 bp in length
* 35999 36098: gap of 100 bp
* 36099 36781: contig of 683 bp in length
* 36782 36881: gap of 100 bp
* 36882 37601: contig of 720 bp in length
* 37602 37701: gap of 100 bp
* 37702 38409: contig of 708 bp in length
* 38410 38509: gap of 100 bp
* 38510 39210: contig of 701 bp in length
* 39211 39310: gap of 100 bp
* 39311 40022: contig of 712 bp in length
* 40023 40122: gap of 100 bp
* 40123 40834: contig of 712 bp in length
* 40835 40934: gap of 100 bp
* 40935 41660: contig of 726 bp in length
* 41661 41760: gap of 100 bp

TITLE
JOURNAL
COMMENT

[illegible]

PUBMED FEATURES	2246248	Location/Qualifiers			
		1. .3327	/organism="Homo sapiens"	/db_xref="taxon:9606"	
source		/map="19p13.1"	/clone="pHMP2"	1. .3327	
		/gene="CRTM"	<1. .1023	/gene="CRTM"	
gene		/codon_start=1	/evidence=experimental	/product="cartilage matrix protein"	
		/protein_id="AAAG3904.1"	/db_xref="GI:180654"	/db_xref="GDB:G00-127-280"	
CDS		/translation="PQDSVDVSARARASGVLEFAIGVSGVDKATLRQIASERQDEHV DYVESYVIEKLSKRKFOEAFVCSVDLCAQGDHCEQVCISSPGSYTCACHEGFTLNSD GKTCNVSGGGSSATDLVFLIDGSKSVRPENLELVKFIQSIVDTLDVSDKLAQVGL VOYSSSVRQEPFLGRFHTKDKIAAVRNMSYMERGTMTGAALKYLIDNSTVSSGARP GAOKVGIVFTDGRSODYINDAKKAKDLGFKMFAGVGVGNAVEDELREIASEPVAEHVF YTADFRTINGIKRLQKKICVEEDPCAESLVKFAQKVEGLLQALTRKLEAVSKRLAI LENTV"			
		<1. .196	/gene="CRTM"	/note="G00-127-280"	
exon		197. .322	/number=3	/gene="CRTM"	
		/note="G00-127-280"	/number=4	/gene="CRTM"	
exon		323. .739	/number=5	/gene="CRTM"	
		/note="G00-127-280"	/number=5	/gene="CRTM"	
exon		740. .892	/number=6	/gene="CRTM"	
		/note="G00-127-280"	/number=6	/gene="CRTM"	
exon		893. .973	/number=7	/gene="CRTM"	
		/note="G00-127-280"	/number=7	/gene="CRTM"	
exon		974. .>3327	/number=8	/gene="CRTM"	
		/note="G00-127-280"	/number=8	/gene="CRTM"	
BASE COUNT	769 a 823 c 949 g 786 t				
		ORIGIN			
Query Match 2.8%; Score 95.6; DB 9; Length 3327;					
Best Local Similarity 50.4%; Pred. No. 8.3e-10;					
Matches 264; Conservative 0; Mismatches 254; Indels 6; Gaps 1;					
Oy	422	AAACCATCGGAAGATTTACGTCGCCACAAATGATGTGTGTCGCTCGACGGGACA 481			
Db	299	ACAGCGACGCGCAAGACCTGCAATGCTGACGTGGTGGTGGCGACCTCGGCCACTGACC 358			
Oy	482	TCATGTTTCTGTAGATGGTCTACACGCTCGGGAAGGAGGCTTTGAAAGGTCACAGC 541			
Db	359	TGGTCTTCTCATTGACGGATCCAAAGAGTGTGAGGCCAGAGAACCTTGAGCTGGTGAAGA 418			
Oy	542	ACTTTGCCATCACAGTCTGTGACGCTCTGGACATCAGCCCGAGAGGCTCAGAGTGGAG 601			
Db	419	AGTTCATCAGATCTGTGATGCTGGAGCTGTGACAGACGCTGCGCCAGGTGGGGC 478			
Oy	602	CATTCAGTTCAGTTCCTCCTCATCTGGAATTCCTCCCTTGGATTCTATTTCAACCCCAAC 661			
Db	479	TGGTGCAGTACTCAAGCTCTGTGCGCCAGGAGTTCCCTCCCTGGTGGTCTCCACACCAAGA 538			
Oy	662	AGGAAGTGAAGCGAAGATCAAGAGGATGGTTTTCAAAGGAGGCGGCACGGAGCGGAAC 721			

Db	539	AGGACATCAAGCGCGGTGTGCGGAATATGTCTCTACATCGAGGAAGCGGCACAAATGA .CGGGG	598		
Qy	722	TTGCTCTGAATACCTTCTGCACAGAGGGTTCCTCGGAGGAGCAGAAATGCT-----TCTG	775		
Db	599	CTGCTCTCAAGTACCTCATTGACAATTCCTTCACTGTGTCCAGTGGGCTAGGCCCGGGG	658		
Qy	776	TGCCCCAGATCTCATCTGCTCACTGATGGGAAGTCCACAGGGGATGTGCACACTGCCAT	835		
Db	659	CCCAAGAGGTGGGCATTTCTTCACTGATGCGCCGAGCAGGACTACATTAAATGATGCTG	718		
Qy	836	CCAAGCAGCTGAAGGAAAGGGGTGCTCACTGTGTTTGTGCTGGGGTCAAGTTTCCACAGT	895		
Db	719	CAAGAAGGCCAAGACCTCGGCTTTAAGATGTTTGTGTGGTGTGGCAATGCCGTGG	778		
Qy	896	GGGAGGAGCTGCATGCTGCGCAGCAGCCTAGAGGGCAGCAC	939		
Db	779	AGGATGAGCTGAGGGAATAGCTCAGAACCTGTGGCAGAGCAC	822		
RESULT 12					
AB040943					
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
Homo sapiens CDNA to mRNA, clone_lib:pBluescriptII SK plus					
clone:fg02934.					
ORGANISM					
Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE					
AUTHORS					
TITLE					
Nagase,T., Kikuno,R., Ishikawa,K., Hirotsawa,M. and Ohara,O.					
Prediction of the coding sequences of unidentified human genes.					
XVII. The complete sequences of 100 new cDNA clones from brain					
which code for large proteins in vitro					
DNA Res. 7 (2), 143-150 (2000)					
JOURNAL					
MEDLINE					
REFERENCE					
2 (bases 1 to 7504)					
Ohara,O., Nagase,T. and Kikuno,R.					
Direct Submission					
TITLE					
Submitted (04-APR-2000) Osamu Ohara, Kazusa DNA Research Institute,					
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba					
292-0812, Japan (E-mail:cdna@info.kazusa.or.jp,					
URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,					
Fax:81-438-52-3914)					
FEATURES					
Location/Qualifiers					
1. .7504					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone="fg02934"					
/clone_lib="pbluescriptII SK plus"					
1. .7504					
/gene="KIAA1510"					
<1. .3425					
/gene="KIAA1510"					
/note="Start codon is not identified. fg02934 cDNA clone					
for KIAA1510 has 286-bp, 143-bp and 96-bp insertions after					
the positions 2358, 3078 and 3274 of the sequence of					
KIAA1510, respectively."					
/codon_start=3					
/product="KIAA1510 protein"					
/protein_id="BAA96034.1"					
/db_xref="GI:7959281"					
/translation="PSQDPRTPGSEWRETGPQRCPLPPVADMVFLVDSWSIGHSIH					
FQQVKDFLASVIAPFEIPDKVQVGLTQYSGDAQTEMDLNSLSTKQVLAARLRRLIK					
GGNTFTGLATHVLGQNLQPAAGLRPAKVVILVTDGKSQDDVHTAARVLDGLVNV					
FAGVYKNADEALRLASPPRDTITVPSHAFVLDLQALGALLSLR.LICORLAGGSPGPG					
AAAPALDTPAPTSLSVLVSQVSSIRSLWTPAPRPLKYLIVRASRGTPPREVYVH					
PAASELHNLASRTEYLVSVPPIYEGGVGEGRLGVTAPLPAPRALTLAAVTPRVH					
LTWPSAGATHLVRCSPASPGEEEREVOVGRPEVLIDGLEPGRDYEVSQSLRG					
EGSEARGIRARTPTLAPPHHIGFSDVSHDAARFWEAGPRVLRVRYTVSSSGHSH					
QTEAPGNATSMGLPLSSSTTYTVRVTCLYPGGSGSTLTGRVTTKKAPSPSLSMTEL					

Db	539	AGGACATCAAGCGGCTGTGCGGAATATGTCTCTACATGGAAGAGGCCACATGA..CGGGG 598			
Oy	722	TTGCTCTGAAATACCTTCTGCACAGAGGCTTGCCTGGAGGAGCAAAATGCT-----TCTG 775			
Db	599	CTGCTCTCAAGTACTCATTCACAAATCCCTTCATCTGTCTCAGTGGGCTAGGCCGGGG 658			
Oy	776	TGCCCCAGATCCTCATCATCTACTGATGGGAAGTCCAGGGGATGTCGACACTGCCAT 835			
Db	659	CCAGAAGGTGGCATTTGCTTCTACTGATGCCGAGCAGGACTACATTAATGATGCTG 718			
Oy	836	CCAAGCAGCTGAAGAAAGGGGTGTCACCTGTGTTTGTCTGGGGTCAAGTTTCCACAGT 895			
Db	719	CCAAGAAGGCCAAAGACCTCGGCTTTAAGATGTTTGTCTGGGTGGGCAATGCCGTGG 778			
Oy	896	GGGAGGAGCTGCATGCACCTGCCAGCAGCCTAGAGGGCAGCAC 939			
Db	779	AGGATGAGCTCAGGAAATACCTCAGAACCTGTGGCAGAGCAC 822			
RESULT 12					
AB040943					
LOCUS		7504 bp mRNA linear PRI 22-FEB-2001			
DEFINITION		Homo sapiens mRNA for KIAA1510 protein, partial cds.			
ACCESSION		AB040943			
VERSION		AB040943.1 GI:7959280			
KEYWORDS		Homo sapiens cDNA to mRNA, clone_lib:pBluescriptII SK plus			
SOURCE		clone:fg02934.			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
TITLE		Nagase,T., Kikuno,R., Ishikawa,K., Hirose,M., and Ohara,O.			
JOURNAL		Prediction of the coding sequences of unidentified human genes.			
MEDLINE		XVII. The complete sequences of 100 new cDNA clones from brain			
REFERENCE		which code for large proteins in vitro			
AUTHORS		DNA Res. 7 (2), 143-150 (2000)			
TITLE		20277482			
JOURNAL		2 (bases 1 to 7504)			
MEDLINE		Ohara,O., Nagase,T. and Kikuno,R.			
REFERENCE		Direct Submission			
AUTHORS		Submitted (04-APR-2000) Osamu Ohara, Kazusa DNA Research Institute,			
TITLE		Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba			
JOURNAL		292-0812, Japan (E-mail:cdna@kazusa.or.jp,			
FEATURES		URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,			
source		Fax:81-438-52-3914)			
Location/Qualifiers					
1. .7504		/organism="Homo sapiens"			
/db_xref="taxon:9606"		/clone="fg02934"			
/clone_lib="pBluescriptII SK plus"		1. .7504			
/gene="KIAA1510"		<1. .3425			
/note="Start codon is not identified. fg02934 cDNA clone		for KIAA1510 has 286-bp, 143-bp and 96-bp insertions after			
the positions 2358, 3078 and 3274 of the sequence of		KIAA1510, respectively."			
/codon_start=3		/product="KIAA1510 protein"			
/protein_id="BAA96034.1"		/db_xref="GI:7959281"			
/translation="PSODPRTGSGSEHRETGPORCLPPVPADAVFLVDGWSIGHSH		FQVKDFLASVIAPFEIGDPKQVQLVLTQYSGDAQOTENDLSLSTKEQVLAARLRK			
GGNTFTGLTHVLGQNLQPAAGLRPAKAVVILVTDGKSODDVHTAARVLKDLGVNV		FAYGVKNDAEALRLASPPRDTIVHSVLDLFQLQALAGLLSR..ICQRLAGGSPROG			
AAAPALDTLPAPTSLVLSQVTSSTSSIRLSWTAPRHLKYLTVRASRGGTPREVYVGH		PAASTELNLASRTEYLVSVEPIYEGVGEGELGLVTTAPLPAPRALTAAPTPTVTVH			
LTWQPSAGATHYLVRCSPASPKGEEEREQVQVRPEVLLUGLEPRDEYEVSVQSLRGP		EGSEARGNARTPTLAPRHLGFSVSDHDAARFWEAGAPRVRVLRVTVYSSEGGHSG			
QTEAPGNATSMILGPLSSSTTYTVRVTVTCLYPGGSSLTGRVTTKAPSPQSLSMTEL					

PGDAVOLAWAAAPSGVLVYQITWTPGLGCKAHEIISVPGNLGTAVLPGLGRHTEYDVT
ILAYYRDGARSQDVSLRYTPTSTVSPPSNALASETPSLOQVSTWTPGLGRVLYHMLT
YAPASGLPKESYVPGCARSHVTLPLDIAATKIRVLVYSALYAAAGREASVATGQATCP
ALRPDGLPGFDLWAFSLVEKAYASIRGVAMEPSAFGGTPTTFLPKDAOLTRVSDV
YPAPLEPHITVFLVPLTPTPEAFALQWMTAEDQPPLGLVLLDAGKKSLLTYFHRDP
RAALQAEATDPQEVRLIFFGTFPHKVVHVAQSRKVLIVDQCKRAVLRBGMGSPPAAP
FVTLGRALKARGPRSSAQLQMLQITVCSDTWADERCCLEPASDRGTCFPAVSAC
SCSETPPGPGPGPGLPCRNCTPGEEQPGDPGPPGKGEKDHLGLPCLOGPFGHQ
GIPGRVLOGPKGMRLGCTAGLPGPPGPGFGOMAGACTSGERGPPGTVGTGLPG
PKGEREKEBQSLATLYQLVQCSASAQTQVHSKFDSEFHENTRPMPLILEQKLEPCT
EPUGSREKSPQVLEWGRGHRLEGRGEPGAVGQMGSPGQCGASTQGLWE"

1530 a 2336 c 2283 g 1355 t

BASE COUNT
ORIGIN

[illegible]

RESULT 13				
AC112730/c				
LOCUS	AC112730	194143 bp	DNA	linear
DEFINITION	Rattus norvegicus clone CH230-8B11, *** SEQUENCING IN PROGRESS ***, 70 unordered pieces.			
				HTG 10-JUL-2002

AC112730
VERSION AC112730.2 GI:21716976
KEYWORDS HTG: HTGS PHASE1.

SOURCE. Rattus norvegicus
ORGANISM. Norway rat.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE. 1 (bases 1 to 194143).

1 (bases 1 to 194143)

REFERENCE
AUTHORS

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.B., Amaratunga, H.C., Are, J.R., Ayie, M., Banks, T., Barbara, J., Benton, J., Blmage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,

Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Benn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Haviak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B.,
 Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Kelly, S., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kell, Y., S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseged, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczkyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, J.,
 Weinstein, G. and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 194143)
 Worley, K.C.

Direct Submission
 Submitted (24-FEB-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 194143)
 Worley, K.C.

Direct Submission
 Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 9, 2002 this sequence version replaced gi:18874114.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GCPA
 Center clone name: CH230-8B11
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap; version 0.930329
 Consensus quality: 131293 bases at least Q40
 Consensus quality: 137579 bases at least Q30
 Consensus quality: 142936 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 70 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

*	1	1071:	contig of 1071	bp in length
*	1072	1171:	gap of unknown length	
*	1172	2607:	contig of 1436	bp in length
*	2608	2707:	gap of unknown length	
*	2708	3724:	contig of 1017	bp in length
*	3725	3824:	gap of unknown length	
*	3825	4860:	contig of 1036	bp in length
*	4861	4960:	gap of unknown length	
*	4961	6734:	contig of 1774	bp in length
*	6735	6834:	gap of unknown length	
*	6835	8334:	contig of 1503	bp in length
*	8335	8434:	gap of unknown length	
*	8435	9763:	contig of 1329	bp in length
*	9764	9863:	gap of unknown length	
*	9864	11408:	contig of 1545	bp in length
*	11409	11508:	gap of unknown length	
*	11509	12933:	contig of 1425	bp in length
*	12934	13033:	gap of unknown length	
*	13034	14051:	contig of 1018	bp in length
*	14052	14151:	gap of unknown length	
*	14152	15865:	contig of 1714	bp in length
*	15866	15965:	gap of unknown length	
*	15966	17350:	contig of 1385	bp in length
*	17351	17450:	gap of unknown length	
*	17451	18750:	contig of 1300	bp in length
*	18751	18850:	gap of unknown length	
*	18851	20131:	contig of 1281	bp in length
*	20132	20231:	gap of unknown length	
*	20232	21916:	contig of 1685	bp in length
*	21917	22016:	gap of unknown length	
*	22017	23459:	contig of 1443	bp in length
*	23460	23559:	gap of unknown length	
*	23560	25079:	contig of 1520	bp in length
*	25080	25179:	gap of unknown length	
*	25180	26233:	contig of 1054	bp in length
*	26234	26333:	gap of unknown length	
*	26334	27837:	contig of 1504	bp in length
*	27838	27937:	gap of unknown length	
*	27938	29298:	contig of 1361	bp in length
*	29299	29398:	gap of unknown length	
*	29399	30500:	contig of 1102	bp in length
*	30501	36000:	gap of unknown length	
*	36001	32066:	contig of 1466	bp in length
*	32067	32166:	gap of unknown length	
*	32167	34285:	contig of 2119	bp in length
*	34286	34385:	gap of unknown length	
*	34386	37679:	contig of 1384	bp in length
*	37679	38669:	gap of unknown length	
*	38670	37889:	contig of 2020	bp in length
*	37890	37989:	gap of unknown length	
*	37990	39137:	contig of 1148	bp in length
*	39138	39237:	gap of unknown length	
*	39238	40913:	contig of 1676	bp in length
*	40914	41013:	gap of unknown length	
*	41014	42688:	contig of 1675	bp in length
*	42689	42788:	gap of unknown length	
*	42789	44673:	contig of 1885	bp in length
*	44674	44773:	gap of unknown length	
*	44774	45816:	contig of 1043	bp in length
*	45817	45916:	gap of unknown length	
*	45917	47849:	contig of 1933	bp in length
*	47850	47949:	gap of unknown length	
*	47950	49720:	contig of 1771	bp in length
*	49721	49820:	gap of unknown length	
*	49821	51279:	contig of 1459	bp in length
*	51280	51379:	gap of unknown length	
*	51380	53053:	contig of 1674	bp in length
*	53054	53153:	gap of unknown length	
*	53154	55373:	contig of 2220	bp in length
*	55374	55473:	gap of unknown length	
*	55474	56590:	contig of 1117	bp in length
*	56591	56690:	gap of unknown length	
*	56691	58480:	contig of 1790	bp in length

*	58481	58580:	gap of	unknown	length
*	58581	61198:	contig of	2618	bp in length
*	61199	61298:	gap of	unknown	length
*	61299	63303:	contig of	2005	bp in length
*	63304	63403:	gap of	unknown	length
*	63404	65179:	contig of	1776	bp in length
*	65180	65279:	gap of	unknown	length
*	65280	68386:	contig of	3107	bp in length
*	68387	68486:	gap of	unknown	length
*	68487	70695:	contig of	2209	bp in length
*	70696	70795:	gap of	unknown	length
*	70796	74237:	contig of	3442	bp in length
*	74238	74337:	gap of	unknown	length
*	74338	77340:	contig of	3003	bp in length
*	77341	77440:	gap of	unknown	length
*	77441	80628:	contig of	3188	bp in length
*	80629	80728:	gap of	unknown	length
*	80729	83845:	contig of	3117	bp in length
*	83846	83945:	gap of	unknown	length
*	83946	86807:	contig of	2862	bp in length
*	86808	86907:	gap of	unknown	length
*	86908	90318:	contig of	3411	bp in length
*	90319	90418:	gap of	unknown	length
*	90419	93228:	contig of	2810	bp in length
*	93229	93328:	gap of	unknown	length
*	93329	95642:	contig of	2314	bp in length
*	95643	95742:	gap of	unknown	length
*	95743	98304:	contig of	2562	bp in length
*	98305	98404:	gap of	unknown	length
*	98405	103325:	contig of	4921	bp in length
*	103326	103425:	gap of	unknown	length
*	103426	105503:	contig of	3078	bp in length
*	105504	106603:	gap of	unknown	length
*	106604	110084:	contig of	3481	bp in length

Query Match	2.6%;	Score 86.8;	DB 2;	Length 194143;
Best Local Similarity	53.6%;	Pred. NO. 5e-08;		
Matches 181;	Conservative	0;	Mismatches 157;	Indels 0;
				Gaps 0;

Qy 1893 GCCAGGCTGCCGAGACAAAGCCCTTGGACCTCGTCTTCATGTTTGGACACACTCTGCGCTCACT 1952

Db 118487 GCGAGGCTGCCGAGGGCCACCGTGGATCTGTCTCTCGTGGATGGCTCCAA*AGTGT 118428

Qy 1953 AGGCCCGCAGAA*TTTTCCTCAGATGCAGAGACTTTGTGCAGAGCTGTGCCCTCCAGTTTCA 2012

Db 118427 GCGTCCACAGAACTTCAGAGCTGGTGAAGCGCTTCGTGAACCAAGATTGTGGAT*CTCTCGA 118368

Qy 2013 GGTCAACCCCTACGTGACACAGCTCGCCCTGGTGGTATGTCGAGCAGCCAGTGCAGACTGC 2072

Db 118367 CGTGTCCCCCTGAGGGCACACGTGTGGGCTGGTGGCATGTTCTTAGCCCCGGTGGCGACTGA 118308

Qy 2073 CTTCCGGCTGACACCAACCCACCCGGGCTGGCATGCTGCGGGCCATTAGCCAGAGCCCC 2132

Db 118307 GTTCCCTCTGGCCGCTATGGCACCCGGCTGAGGTGAAGCAGCGAGTTTGGCCGCTCGA 118248

Qy 2133 CTACCTAGGTGGGTGGGCTCAGCCGCAACCGCCCTGCTGCACATCTATGACAAATGAT 2192

Db 118247 GTACATGGCGCGGCACCATGACAGGCTGGCACTGCGTCACATGTTGGAGCACAGTTT 118188

Qy 2193 GACCGTCCAGAGGGTGCCCGGCTGGTGTCCCCAAAG 2230

Db 118187 CTCAGAGGTGCAGGGCGCGGGCTCGAGACCTCAACG 118150

RESULT 14

MMU35035

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS
SOURCE

SOURCE

MMU35035 159 bp mRNA linear ROD 23-AUG-1996
Mus musculus cartilage matrix protein precursor (CMP) mRNA, partial
cds.
U35035
U35035.1 GI:1163178
house mouse strain-C57 Black/6 x CBA.

```

ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
               1 (bases 1 to 1959)
TITLE          Aszodi,A., Hauser,N., Studer,D., Paulsson,M., Hiripi,L. and
               Bosze,Z.
               Cloning, sequencing and expression analysis of mouse cartilage
               matrix protein cDNA
JOURNAL        Eur. J. Biochem. 236 (3), 970-977 (1996)
MEDLINE        96270751
PUBMED        8665920
REFERENCE      2 (bases 1 to 1959)
AUTHORS        Aszodi,A.
TITLE          Direct Submission
JOURNAL        Submitted (30-AUG-1995) Attila Aszodi, Agricultural Biotechnology
               Center, Institut for Animal Sciences, Szent-Gyorgyi Albert 4.,
               Godollo H-2101, Hungary
COMMENT        On Jan 23, 1996 this sequence version replaced gi:1144533.
FEATURES       Location/Qualifiers
               1..1959
               /organism="Mus musculus"
               /strain="C57 Black/6 X CBA"
               /db_xref="taxon:10090"
               /clone="mCMP"
               /tissue_type="cartilage"
               /dev_stage="adult"
               /lab_host="Mus musculus"
               1..1959
               /gene="CMP"
               1..1959
               /gene="CMP"
               /product="cartilage matrix protein precursor"
               /note="putative"
               1..1503
               /gene="CMP"
               /note="putative"
               /codon_start=1
               /product="cartilage matrix protein precursor"
               /protein_id="AAB0521.1"
               /db_xref="GI:1163179"
               /translation="MKVTSGPSALCSLLLLLLLQVPSLSLPQPHLCRTPT
               DSVFVDSRSVPVEFKVFLQVIESLDVGNATRVGLVNVASVVKPEFPLRAH
               GSKALQARRIOPLSTGTMFLQALQFAITKALSDAEGGRARSPDISKVIWTDGR
               PQDSVRDYSERARASGIELEAIGLGRVDKATLROIASEPQEDVDVSYNVIEKLAK
               KFEQAFVYSDLCATGDHCEQLCVSSPGSTCACHGEFTLNSDKTNCVRGSGSGS
               ATDLFLDGSKVRPENFELVKFQINQIVDTLDVSDRLAQVLQVYSSSIRQEPPLG
               RFHKKDIKARVRNMSYMEKGTMTGAALKYLDNSFTVSSGARPGAQKVGFVTDGRS
               QDIINDAARKADLGFKMFVGVGNVAVEELREIASEPVAHFYTFADFKTINQIGKK
               LQKQICVEDPCACESILKFEAKVEGLLQALTRKLEAVSGRLAVLENRII"
               1..87
               sig_peptide
               /gene="CMP"
               /note="putative"
               88..1500
               /gene="CMP"
               /product="cartilage matrix protein"
               /note="putative"
               1504..1959
               /gene="CMP"
               403 a 501 c 600 g 455 t
BASE COUNT    403 a 501 c 600 g 455 t
ORIGIN
Query Match      2.5%; Score 85; DB 10; Length 1959;
Best Local Similarity 49.2%; Pred. No. 1.5e-07;
Matches 223; Conservative 0; Mismatches 230; Indels 0; Gaps 0;
QY 1911 AGCCCTGGACCTGCTTTCATGTTGGACACCTCTGCTCAGTAGGCGCCGAGAAATTTTGC 1970
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 828 AGCCACCGACCTGGTCTTCTCATCGACGATCCACAGACGCTGCGGCTTGAGACTTTGA 887
QY 1971 TCAGATGCAGACCTTTGTGAGAACGTGTGCGCTCCAGTTTGTAGGTGAACCTGACGTGAC 2030
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 888 GCTGGTGAAGAAGTTTCATCAACACAGATTGTGGACACGTTAGATGTGTCGGACAGGCTAGC 947

```

```

QY 2031 ACAGGTGGCCCTGGTGTATGGCGAGCCAGGTGCAGACTGCTTCGGGCTGCACACCAA 2090
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 948 CCAGGTGGGCTGGTGCAGTACTCTCCAGTCCACTCCAGGAGTTCCTCCACTCGGCCGCT 1007
QY 2091 ACCACCCGGCTGCGATGCTGCGGGCCATTAGCCAGGCCCCCTACCTAGGTGGGTGGG 2150
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1008 CCATCCAGAGAGGACATTAAGCGCGGTGGGACATGTCTTACATGGAGAAAGGCAC 1067
QY 2151 CTCAGCGCGCACCGCCTGCTGCACATCTATGACAAAGTGATGACCGCTCCAGAGSGGTG 2210
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1068 CATGACTGCGCGCCTTGAAGTATCTCATAGATAATTTCTTCACTGTGTCCAGCGGGC 1127
QY 2211 CCGGCTGTGTGTCCTCCCAAGCTGTGTGTGTCACAGCGGGGAGAGGCGCAGAGATGC 2270
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1128 AAGGCTTGAGGCCACAGAAGTGGCATCTCTTCCAGCATGGCCGAGGACGATACAT 1187
QY 2271 AGCGGTCTCTGCCACAGCTGAGGAACAATGCATCTCTGTCTGTGTCGTGGCGCTGGG 2330
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1188 TAATGACGTGCCAGGAGGCCAAGGACCTTGCTTTAAGATGTTTTCGGTGGCGGTGGG 1247
QY 2331 GCCTGTCTTAAGTGGGTCTGCGAGGCTTGC 2363
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1248 CAATGCTGTGGAGAGAGCTGAGGAGATCGC 1280

RESULT 15
AL590429      Mouse DNA sequence from clone RP23-462016 on chromosome 2, complete
LOCUS        sequence.
DEFINITION   AL590429.9 GI:17017748
ACCESSION    AL590429
VERSION      HTG.
KEYWORDS     Mus musculus
SOURCE       house mouse.
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
               1 (bases 1 to 190669)
REFERENCE    1 (bases 1 to 190669)
AUTHORS      Kay,M.
TITLE        Direct Submission
JOURNAL      Submitted (17-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
               Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
               humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
               On Nov 20, 2001 this sequence version replaced gi:15020991.
COMMENT      During sequence assembly data is compared from overlapping clones.
               Where differences are found these are annotated as variations
               together with a note of the overlapping clone name. Note that the
               variation annotation may not be found in the sequence submission
               corresponding to the overlapping clone, as we submit sequences with
               only a small overlap as described above.
               This sequence was finished as follows unless otherwise noted: all
               regions were either double-stranded or sequenced with an alternate
               chemistry or covered by high quality data (i.e., phred quality >=
               30); an attempt was made to resolve all sequencing problems, such
               as compressions and repeats; all regions were covered by at least
               one plasmid subclone or more than one M13 subclone; and the
               assembly was confirmed by restriction digest. The following
               abbreviations are used to associate primary accession numbers given
               in the feature table with their source databases: Em:, EMBL; Sw:,
               SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
               database can be found at
               http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-462016 is
               from the RPI-23 Mouse PAC Library
               constructed by the group of Pieter de Jong.
               For further details see http://www.chori.org/bacpac/home.htm
               VECTOR: pBACE3.6
               This sequence is the entire insert of clone RP23-462016 The true
               left end of clone RP23-14014 is at 187423 in this sequence. The
               true right end of clone RP23-346D16 is at 20063 in this sequence.
               Location/Qualifiers
               1..190669
               /organism="Mus musculus"
               /db_xref="taxon:10090"
FEATURES      source

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 22:43:48 ; Search time 465 Seconds
(without alignments)
16345.149 Million cell updates/sec

Title: US-09-930-020A-1

Perfect score: 3375

Sequence: 1 gacagtgttcgcggctgcac.....tgtccaccttggaaggtcttc 3375

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3375	100.0	3375	24	ABK92207
2	2020.8	59.9	3485	23	AA581530
3	1117.2	33.1	1837	21	AA261803
4	1117.2	33.1	1837	22	AA261803
5	1117.2	33.1	1837	24	AB348888
6	340	10.1	406	23	AA381529
7	201	6.0	651	24	AB016108
8	201	6.0	651	24	AB016109
9	177	5.2	651	24	ABQ16106

c	10	177	5.2	651	24	ABQ16107	Oligonucleotide fo
	11	98.8	2.9	1491	23	AA591851	DNA encoding novel
	12	97.8	2.9	4151	24	AA594773	Human DNA sequence
	13	74.6	2.2	1668	20	AA000664	Human secreted pro
	14	74.2	2.2	782	22	AA003933	Human CDNA clone (
	15	74.2	2.2	1167	22	AAH17154	Human CDNA sequenc
	16	74.2	2.2	1480	22	AAK94095	Human full-length
	17	74.2	2.2	2043	22	AAK94655	Human full-length
	18	74.2	2.2	2128	22	AAK94169	Human full-length
	19	74.2	2.2	2281	22	AAK94167	Human full-length
	20	73.8	2.2	2000	22	AA08394	Human secreted pro
	21	73.6	2.2	1033	24	ABL61927	Colon adenocarcino
c	22	73.6	2.2	2163	21	AA669634	Human secreted pro
	23	73.6	2.2	2293	22	AAH18007	Human CDNA sequenc
	24	73.6	2.2	2863	20	AA241311	Human normal ovari
	25	73.6	2.2	3372	20	AA18924	Human matrilin-3 e
	26	73.6	2.2	3472	22	AAF30192	Clone 14578444.0.4
	27	73.6	2.2	3449	20	AA52220	Protein PRO219 CDN
	28	73.6	2.2	3449	21	AAA49719	Human PRO219 CDNA
	29	73.6	2.2	3449	22	AA521390	Human CDNA sequenc
	30	73.6	2.2	3449	22	AAF72378	Human PRO219 CDNA
	31	73.6	2.2	3449	22	AA67417	Human angiogenesis
	32	73.6	2.2	3496	24	ABK64744	Human benign prost
	33	73.6	2.2	3711	22	AA08308	Human secreted pro
	34	73	2.2	729	24	ABN86920	Human NOV7 encodin
	35	72.6	2.2	1232	21	AA673867	Human secreted pro
	36	72.6	2.2	3026	22	AAF30191	Clone 14578444.0.1
	37	70.6	2.1	2366	22	AA693767	Human CDNA encodin
	38	69	2.0	9192	24	ABL92117	Human Tumour Endot
	39	69	2.0	9483	22	AA544690	Human full-length
	40	69	2.0	11657	23	ABV23446	Human prostate exp
	41	69	2.0	11657	23	ABV29303	Human prostate exp
	42	68.2	2.0	987	22	AA684021	Human zwf1 polype
	43	68.2	2.0	987	22	AA684021	Human zwf1 gene d
	44	67.8	2.0	898	21	AA676790	Human OREF ORF2345
	45	67.4	2.0	733	22	AA080062	Human extracellular

ALIGNMENTS

RESULT 1

ABK92207

ID ABK92207 standard; DNA; 3375 BP.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

Drmanac RT, Liu C, Tang YT;
WPI: 2001-639362/73.
P-PSDB; ABG17343.
New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
Claim 1; SEQ ID No 17334; 103pp; English.
The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.
Note: the sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
Sequence 3485 BP; 829 A; 850 C; 1030 G; 776 T; 0 other;

Query Match 59.9%; Score 2020.8; DB 23; Length 3485;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 2185; Conservative 0; Mismatches 32; Indels 135; Gaps 1;

QY 400 CAGGAAGTCCATGTAAGCAAGAAACCATCGGGAACATTCAGCTCCAGCAAAATGATG 459
DB 1269 CAGGAAGTCCATGTAAGCAAGAAACCATCGGGAACATTCAGCTCCAGCAAAATGATG 1328
QY 460 TGTGCTCGGCTCAGTGGACATCATGTTCTGTTAGATGGGTCTAACAGCGTCGGGAAA 519
DB 1329 TGTGCTCGGCTCAGTGGACATCATGTTCTGTTAGATGGGTCTAACAGCGTCGGGAAA 1388
QY 520 GGGAGCTTTGAAGGTCCAGACATTTGCCATCACAGTCTGTGACGGTCTGGACATCAGC 579
DB 1389 GGGAGCTTTGAAGGTCCAGACATTTGCCATCACAGTCTGTGACGGTCTGGACATCAGC 1448
QY 580 CCGCAGAGGTCTAGTGGAGCATTCACATTCAGTTCCAGTTCACATCTCATCTGGAATTCGCC 639
DB 1449 CCGCAGAGGTCTAGTGGAGCATTCACATTCAGTTCCAGTTCACATCTCATCTGGAATTCGCC 1508
QY 640 TTGGATTATTTTCAACCCCAACAGGAAGTGAAGCAAGAAATCAAGAGATGTTTTCAAA 699
DB 1509 TTGGATTATTTTCAACCCCAACAGGAAGTGAAGCAAGAAATCAAGAGATGTTTTCAAA 1568
QY 700 GGAGGCGCAGGACGGAACCTTGCTGAAATACCTTCTGCACAGAGGTTGCCGTGA 759
DB 1569 GGAGGCGCAGGACGGAACCTTGCTGAAATACCTTCTGCACAGAGGTTGCCGTGA 1628
QY 760 GGCAGAAATGCTTCTGTGCCAGATTCCTCATCTGTCATCTGATGGGAAGTCCAGGGG 819
DB 1629 GGCAGAAATGCTTCTGTGCCAGATTCCTCATCTGTCATCTGATGGGAAGTCCAGGGG 1688
QY 820 GATGTGGCACTGCCATCCAAAGCAGCTGAAGAAAGGGGTGTCACTGTGTGCTGGGG 879
DB 1689 GATGTGGCACTGCCATCCAAAGCAGCTGAAGAAAGGGGTGTCACTGTGTGCTGGGG 1748

QY 880 GTACAGTTTCCAGGTGGAGGAGTGCATGCTAGCTGCGCAGCAGCTAGAGGCGAC 939
DB 1749 GTACAGTTTCCAGGTGGAGGAGTGCATGCTAGCTGCGCAGCAGCTAGAGGCGAC 1808
QY 940 GTGCTTTGGCTAGCAGGTGGAGATGCCAACAGGCTCTTTCAGCAGCCCTCAGCAGC 999
DB 1809 GTGCTTTGGCTAGCAGGTGGAGATGCCAACAGGCTCTTTCAGCAGCCCTCAGCAGC 1868
QY 1000 TCGGCGCATCTGCTCCAGCGCCAGCCAGACTGCAGGCTCGAGGCTCACCCTCTGAGCAGC 1059
DB 1869 TCGGCGCATCTGCTCCAGCGCCAGCCAGCTGGAGGCGCCGAGCTTCTTTCATGGAGCGG 1928
QY 1060 AGGACGCTGGAGATGCTCCGGGAGTTTCGTGGCAATGCCCAATGCTGTGAGAGATGCGCG 1119
DB 1929 TTAATG----- 1934
QY 1120 CGGACCTTTGGGTGCTGGCTGCACACTGTCCCTTCTACAGTGGAGAGAGTGTTCCTA 1179
DB 1935 ----- 1934
QY 1180 ACCCACCCTGCCACCTGCTACAGAGCACCTGCGCCAGGCCCCCTGTGACTCGCAGCCCTGC 1239
DB 1935 -----GGCATCTCTCTGATAGGCCCTGTGACTCGCAGCCCTGC 1973
QY 1240 CAGATGGAGGCACATGTGTTCCAGAAAGACTGGAGCGGTACCAGTGCCTTCTGCCCGCTG 1299
DB 1974 CAGATGGAGGCACATGTGTTCCAGAAAGACTGGAGCGGTACCAGTGCCTTCTGCCCGCTG 2033
QY 1300 GCCTTTGGAGGGAGGCTAACTGTGCCCTGAAGCTGAGCTTGGAAATGCAAGGTCGACCTC 1359
DB 2034 GCCTTTGGAGGGAGGCTAACTGTGCCCTGAAGCTGAGCTTGGAAATGCAAGGTCGACCTC 2093
QY 1360 CTCTTCTCTGTCGACAGCTCTCGGGCACACTCTCTGGAGCGGTCTCTCGGGGGCCAAAGTC 1419
DB 2094 CTCTTCTCTGTCGACAGCTCTCGGGCACACTCTCTGGAGCGGTCTCTCGGGGGCCAAAGTC 2153
QY 1420 TTGCTGAAGCGGTTTGTGCGGCGGTGCTGAGGAGGACTCTCGGGCGCGAGTGGGTGTG 1479
DB 2154 TTGCTGAAGCGGTTTGTGCGGCGGTGCTGAGGAGGACTCTCGGGCGCGAGTGGGTGTG 2213
QY 1480 GCCACATACAGCAGGAGCTGCTGGGCGGTGCTGCGGGAGTACACAGGATGTCCT 1539
DB 2214 GCCACATACAGCAGGAGCTGCTGGGCGGTGCTGCGGGAGTACACAGGATGTCCT 2273
QY 1540 GACCTGCTTGGAGCCTCGATGSCATTCCTTCCTGCTGGCGCCACCTTGCAGGGCAGT 1599
DB 2274 GACCTGCTTGGAGCCTCGATGSCATTCCTTCCTGCTGGCGCCACCTTGCAGGGCAGT 2333
QY 1600 GCCTTGGCGCAGCGCGCAGCGTGGCTTTCGGGAGCGCCACACAGGACAGCCAGGACCGG 1659
DB 2334 GCCTTGGCGCAGCGCGCAGCGTGGCTTTCGGGAGCGCCACACAGGACAGCCAGGACCGG 2393
QY 1660 CCACGTAGAGTGGTGGTGTGCTTGTCTACTGATGACACTCCGAGGATGAGTGGGGGCCCA 1719
DB 2394 CCACGTAGAGTGGTGGTGTGCTTGTCTACTGATGACACTCCGAGGATGAGTGGGGGCCCA 2453
QY 1720 GCGCGTCACCAAGGCGCGAGAGTGTCTCTGCTGGGTGTAGGCGAGTGAAGCGCGTGGCG 1779
DB 2454 GCGCGTCACCAAGGCGCGAGAGTGTCTCTGCTGGGTGTAGGCGAGTGAAGCGCGTGGCG 2513
QY 1780 GCAGAGCTGGAGAGATCACAGCAGCCCAAGAGTGTGATGGTCTACTCGGATCCTCAG 1839
DB 2514 GCAGAGCTGGAGAGATCACAGCAGCCCAAGAGTGTGATGGTCTACTCGGATCCTCAG 2573
QY 1840 GATCTTCAACCAATCCCTGAGCTGCGAGGGAAGCTGTGACCGCGGAGCGCGCCAGGG 1899
DB 2574 GATCTTCAACCAATCCCTGAGCTGCGAGGGAAGCTGTGACCGCGGAGCGCGCCAGGG 2633
QY 1900 TCGCGGACCAAGCCCTGGACCTGCTTTCATGTTGGACACCTTCTGCCCTCAGTAGGGCCC 1959
DB 2634 TCGCGGACCAAGCCCTGGACCTGCTTTCATGTTGGACACCTTCTGCCCTCAGTAGGGCCC 2693
QY 1960 GAGAAATTTTCTCAGATGAGAGCTTTGTGAGAAAGCTGTGCCCTCCAGTTTGGAGTGAAC 2019

QY 722 TTGCTCTGAATACCTTTCACAGAGAGGTTCCTGGAGGCGAGAAATGCTTCTGTGCCCC 781
DB 521 TAGCCCTGAAA--CCCTGTAGCAGAGAGGTTCCTCCGGAGGCGAGAAATGCTTCTGTGCCCC 577
QY 782 AGATCTCATCATCTGTCACCTGAGGAACTCCACAGGGGATGCGACATGCCATCAAGC 841
DB 578 AGATTCTTATATGTCGACGGATGGCAAGTCCACAGGGGCCCTGCGCTCTCCCGGCTAAGC 637
QY 842 AGCTGAAGAAAGGGGTGTCACCTGTGTTTCTGTGGGGTTCAGGTTTCCACAGTGGGAGG 901
DB 638 AGCTGAGAGAAAGGGGCATCGTCGTGTTTCCGTAGGAGTCCGTTTCCAGGTGGAGC 697
QY 902 AGCTGATGCATGCGCCACGCGCTAGAGGGAGCACACATGCTGTGTTGGCTGAGCAGGTGG 961
DB 698 AGCTGCTCATCGTGGCCAGTGAAGCCAGGACCGGATGCTGTGTTGGCTGAGCAAGTGG 757
QY 962 AGGATGCCACCAAGCCCTTTCAGCACCTCAGCAGCTCGGCCATCTCTCCAGGGCCA 1021
DB 758 AGATGCCACCAATAGCCCTTCCAGCACCTCAGCAGCTCGGCCATCTCTCCAGCACCTGCTG 817
QY 1022 CGCCAGACTGCAGGGTCGAGGCTCACCCCTGTGAGCACAGGACGCTGGAGATGGTCCGGG 1081
DB 818 ATCCAGACTGCAGGGTGAACCTCATCCCTGTGAGCGGAGGACGCTGGAGACCGTCAGGG 877
QY 1082 AGTTCCTGCAATGCCCATGCTGAGAGAGATCGCGGGGACCCCTTGGCGTGTGGCTG 1141
DB 878 AGCTGCTGCAATGCTTGTGCTGAGAGAGATCAAGGCAAGCAGACACTGTGCTGGCTC 937
QY 1142 CACACTGTCCCTTCTACAGTGAAGAGAGTGTTCCTAACCCACCTGCCACCTGCTAC 1201
DB 938 TGCCCTGTCCCTTCTACAGTGAAGAGAGTGTTCAGACACACCCCTGCCAATGCTGTA 997
QY 1202 GGACCACTGCCAGGCCCTGTGACTCGACGCCCTGCCAGAAATGAGGACATGTGTTTC 1261
DB 998 GAACCATCTGTCCAGGCCCTGTGACTCCAGCCCTGCCAAATGAGGACGACGTGCATTC 1057
QY 1262 CAGAAGACTGGACGGCTACCACTGCTCTGCCCGTGGCCCTTGGAGGGAGGCTAACT 1321
DB 1058 CAGAAGTGTGGAATAGGTACCACTGTCTGTGCCCACTTGGCAGGGGGAAGTCAACT 1117
QY 1322 GTGCCCTGAAGCTGAGCCTGGAAATGAGGTCAGCCTCTCTCTCTCTCTCTCTCTCTCT 1381
DB 1118 GTGCCCGGAAGTGAAGCTGGAAATGAGATGATCTCTCTCTCTCTCTCTCTCTCTCT 1177
QY 1382 CGGGACCACTGTGACAGCGCTTCTCTCGGGGCCAAAGTCTTGTGAAGCGGTTTGTGCGGG 1441
DB 1178 CAGGCACCACTATTGGGGGCTTCCGGAGGGCCAAAGCCCTTGTCAAGCGCTTGTGCGAG 1237
QY 1442 CGTGTGTGAGCGAGGACTCTCGGGCCCGAGTGGTGTGGCCACATACACAGGAGCTCG 1501
DB 1238 CGTGTGTGAGGAGGACTCCCGAGCCCGGCTTGGGATAGCCAGTTATGGCAGGAATCTAA 1297
QY 1502 TGTGTGCGGTG--CCTGTGGGGAGTACCAGGA--TGTGCTGTGACCTGTCTGTGGAGCTCGA 1559
DB 1298 TGTGTGCGGTGCTCTCGGGAGTACAGCATGTGTGCGGACCTGATCAGGAGCTTGA 1357
QY 1560 TGGCATTCCTTCCTGTGTGGCCCCACCTGTAGCGGCACTGCTTCGGCAGCGGCAG 1619
DB 1358 CAGCATTCCTTTCAGCGGTGGCCGACCCCTAACCGGAGTGCCTTCTCCAGGTGGCAGA 1417
QY 1620 GCGTGTCTCGGAGCGCCACAGGACAGCCAGGACCGCCACCTAGTAGTGGTGTGTTT 1679
DB 1418 GCACGCTTGTGGAGTGCAGCAGGAGTGTGTGAGGACAGGCGCCAGCAGAGTAGTAGTCT 1477
QY 1680 GCTCACTGATCACATCCAGGATGAGTGTGGGGCCCGCGCTCAGCGCTCAGCAAGGCGGG 1739
DB 1478 GCTCACTGATCAGCTCCAGGATGAGTGTGTGGCCAGCAGCTCAGCAAGGCGCTCG 1537
QY 1740 AGAGCTGCTCTCTGCTGGGTGTAGGCAAGTGAAGCCGTGCGGGCAGCAGCTGAGGAGATAC 1799
DB 1538 GGAGCTACTCTCTCTGCGGTGGGCAAGTGAAGTCTCTGAGGCGGAGCTGTTGAAGATCAC 1597

QY 1800 AGGAGCCCCAAAGCATGTGATGGTCTACTCTGGATCTCTCAGGATCTGTTCAACCAATCCC 1859
DB 1598 CGGTAGCCCCGAGCATGTGATGGTCCACAGACACCCCTCAGGACCTGT--CAGCCAAATCC 1655
QY 1860 TGAGCTGTCAGGGAAGCTGTGAGCGCGGAGCGGCGGAGGTGCGGACACAAAGCCCTGGA 1919
DB 1656 AGAGTGCAGAGGAGCTATGAGCAGCCAGCCAGCGGCTGCCAGGACACAGTCACTGGA 1715
QY 1920 CCTCGTCTTTCATGTTGGACACCTCTGCCCTCAGTAGGGCCCGGAGAAATTTTGTCTCAGATGA 1979
DB 1716 CTTGCTTCTCTCTG--TGGAGCTCTCTG--CTCTGTGGGACCTGAGAACTTTTGGCCAAATGCA 1773
QY 1980 GAGCTTTGTGAGAGCTGTGCCCTCCAGTTTGAAGTGAACCTGAGCTGACACAGCTCGG 2039
DB 1774 GAGCTTTCATCAGAAATGCAACCTCCGGTTTGTATGTAATCTCTGATGTGACACAAGTTGG 1833
QY 2040 CCGT 2043
DB 1834 CCGT 1837
RESULT 4
AAC99736
ID AAC99736 standard; cDNA; 1837 BP.
XX AAC99736;
AC AAC99736;
XX
DT 08-MAR-2001 (first entry)
XX
DE Skin cell cDNA, SEQ ID NO: 350.
XX
KW Mouse; skin cell; cytostatic; antiinflammatory; anti-HIV;
KW neotropic; neuroprotective; vulnary; immunomodulatory; vaccine;
KW keratinocyte growth stimulation; cancer; angiogenesis inhibition;
KW inflammation; neurological disease; ss.
XX
OS Mus sp.
XX
XX WO200069884-A2.
PN
XX
PD 23-NOV-2000.
XX
XX 15-MAY-2000; 2000WO-NZ00075.
PF
XX
XX 14-MAY-1999; 9905-0312283.
PR
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA
XX Watson JD, Strachan L, Onrust R, Sleeman M, Kumble KD, Murison JG;
PI WPI; 2001-007495/01.
DR
XX
XX New isolated polynucleotide used in the identification of genetic
PT disorders and encoding polypeptides used for treating inflammatory
PT disease, cancer and neurological diseases -
XX
PS Claim 1; Page 272-273; 352pp; English.
XX
XX The present polynucleotide encodes a polypeptide which is expressed in
CC mammalian skin cells. The polypeptide is useful for stimulating
CC keratinocyte growth and motility, inhibiting the growth of cancer cells,
CC modulating angiogenesis, inhibiting angiogenesis and vascularisation of
CC tumours, modulating skin inflammation, stimulating the growth of
CC epithelial cells, inhibiting the binding of human immunodeficiency virus
CC (HIV)-1 to leukocytes, and treating inflammatory disease, cancer and
CC neurological diseases. The polynucleotide can be used as a marker, in
CC the identification of genetic disorders, and for the design of
CC oligonucleotides for examining expression patterns.
XX
SQ Sequence 1837 BP; 370 A; 541 C; 557 G; 369 T; 0 other;
Query Match 33.1%; Score 1117.2; DB 22; Length 1837;
Best Local Similarity 79.8%; Pred. No. 3.9e-267;

Matches 1392; Conservative 0; Mismatches 343; Indels 9; Gaps 6;

QY 302 CGCGCTCTCTCCGCTATATCAACATGCCCCCTTTCTGTGTGCTGGAGCGGTCTGTG 361
Db 101 CGCGGCTCGCACAGTCTTAACAACATGCTCCACTCTCTCTTACCAGCCATCTACA 160
QY 362 TTTTCTCTGTTTTCAGAGTCCGCCATCTCTCCCTCTCCAGGAGTCCATGTAAACAAAG 421
Db 161 TGTCTCTGTTTTCAGAGTCTCCCGACCATCTCTCTTCAAGGAAGTGCATGTGAACCGGG 220
QY 422 AAACCATCGGAAGATTTCAGCTGCCAGCAAAATGATGTGCTCGGCTGCAGTGGACA 481
Db 221 AGACATGGGGAAGATCGCTGTGCCACCAATTAATGTGTCTAGCCCGCGTCAACA 280
QY 482 TCATGTTTCTGTAGATGGCTTAACAGCGTCGGGAAAGGAGCTTTGAAGGTCCAAGC 541
Db 281 TCCTGTTTCTGTAGATGGCTCTCACAGCATCGGAAGGGAGCTTCGAGAGGTCCAAGC 340
QY 542 ACTTTGCCATCAGATCTGTGACGGTCTGGACATCAGCCCCGAGAGGTGAGAGTGGAG 601
Db 341 GCTTGGCCATCGTGTGATGCGCTGGACATCAGCCCTGGCAGGGTTCAGAGTCGGAG 400
QY 602 CATTTCAAGTTTCACTTCCACCTCATCTGGAATTCCTTGGATTCATTTTCAACCCAAC 661
Db 401 CTTTGCAGTTTGGTTCACCTCTCATCTGGAATTCCTTGGACTCTCTCAACTCGAC 460
QY 662 AGGAAGTGAAGGCAAGAAATCAAGAGGATGTTTTTCAAAGGAGCGGCAGCAGCAAGC 721
Db 461 AGGAAGTGAAGGAAAGCATCAAGGGATAGTTTTTCAAAGTGGCGCACAGCAGCGGCC 520
QY 722 TTGCTCTGAATACCTTCTCACAGAGGTTGCTTGAGGCGAGAAATGCTTCTGTGCCCC 781
Db 521 TAGCCCTGAAA---CGCCTCAGGAGAGGGTTCCCGCGAGGAGAGAAATGGCTCTGTGCCCC 577
QY 782 AGATCTCATCATCTGCTCACTGATGGGAAGTCCAGGGGATGTGGCACTCCCATCCAAGC 841
Db 578 AGATCTTATCATCTGTCAGCGATGGCAGTCCAGGGGCCGTGGCTCTCCCGGCTAAGC 637
QY 842 AGCTGAAGGAAGGGGTGTCACCTGTGTGTGCTGTGGGGTTCAGGTTTCCAGGTGGGAGG 901
Db 638 AGCTGAGAGAAAGGGCATCGTGTGTGTGCTGTGGAGTCCGTTTTTCCAGAGTGGGAGC 697
QY 902 AGCTGATGACATGCGCCAGGAGCTAGAGGGGAGCAGTGTGTGTGGCTGAGCAGGTGG 961
Db 698 AGCTGTCTCAGCTGTGGCAGTGAGCCGAAGGAGCCGCGATGTGCTGTGGCTGAGCAAGTGG 757
QY 962 AGGATGCCAACACGGCTCTTCAACACCTCAGCAGCTCGGCGATCTGCTCCAGCGCCA 1021
Db 758 AGGATGCCAACATGGCTCTTCAACACCTCAGCAGCTCGGCACTCTGCACCACTGCTG 817
QY 1022 CGCCAGACTGAGGCTCGAGGCTACCCCTGTGAGCAGAGAGCTGGAGATGGTCCGGG 1081
Db 818 ATCCAGACTGCAGGTTGGAACCTCATCCCTGTGAGCGAGGAGCGCTGGAGACGCTCAGGG 877
QY 1082 AGTTCCGTGCAATGCCCATCTGAGAGAGGATCGCGGCGGACCCCTTGGCGGTGCTGGCTG 1141
Db 878 AGCTCGCTGCAATGGCTTGTGTGGAGAGGATCAAGGCAAGCAGACACTGTGCTGGCTC 937
QY 1142 CACACTCTCCCTTCTACAGCTGGAAGAGTGTTCCTTAACCCACCTGCCACCTGTGTACA 1201
Db 938 TGCCCTGTCCCTTCTACAGCTGGAAGAGTGTTCAGACAGACACCTGCCCACTGTCTACA 997
QY 1202 GGACCACTGCCAGGCGCTGTGACTCGCAGCCCTGCCAGATGAGAGGACATGTGTTC 1261
Db 998 GAACCATCTCTCCAGGCGCTGTGACTTCCAGCCCTGCCAAATGGAGGACCGTGCATTC 1057
QY 1262 CAGAGGACTGGAGCGCTACCATGCTCTGCCGCTGGCTTGGAGGGAGGCTTACT 1321
Db 1058 CAGAAGGTGTGGATGAGTACCACTGTCTCTGCCACTTGGCATTCGGAGGGGAAGTCAACT 1117
QY 1322 GTGCCCTGGAAGCTGAGCTGGAATGCAAGGTTCGACCTCTCTTCTCTGCTGCAGCTCTG 1381
Db 1118 GTGCCCGAAGCTGAGCTGGAATGCAAGATCGATGTCTCTCTCTCTGCTGCAGAGTTCTG 1177

QY 1382 CGGGCACCACTCTGGACGGGTTCTCTGGGGCCAAAAGTCTTCTGTGAAGCGGTTTGTGGGG 1441
Db 1178 CAGGCACCACTATGGGGGGTTCCTGGAGGGCCAAAGCCCTTGTCAAGCGCTTGTGTGCAAG 1237
QY 1442 CCGTGTGAGCAGGAGTCTCTCGGGCCGAGTGGGTGGCCACATACAGCAGGAGCTGC 1501
Db 1238 CCGTGTGAGGAGGAGTCTCCGAGCCCGCTTGGGATAGCCAGTTATGGCAGGAATCTAA 1297
QY 1502 TGTGTGCGGTG-CCTGTGGGGAGTACCAGGA-TGTGCTCTGACCTGGTCTGGAGCCCTCGA 1559
Db 1298 TGTGTGCGGTGCTCTGTGGGGAGTACCAGCATTTGCGGAGCTGATCAGGAGCCTTGA 1357
QY 1560 TGGCATTTCCCTTCCTGTGGTGGCCCCCCTGACGGGAGTGCCTTGGCGGAGCGGCGAGA 1619
Db 1358 CAGCATTTCCCTTCAGCGGTGGCCCGACCTTAACCGGGAGTGCCTTGTCCAGGTGCGAGA 1417
QY 1620 GCCTGCTTCGGGAGCGCCACAGGACAGCCAGGACGGCCAGCTAGAGTGGTGTCTT 1679
Db 1418 GCACGGCTTTGGGAGTGCACAGAGGACTGTGTCAAGACAGGCGCACGAGTAGTGTCT 1477
QY 1680 GCTCACTGAGTCACTCTCCGAGGATGAGGTTCGGGGCCACAGCGCTCACCAAGGGCGG 1739
Db 1478 GCTCACTGATCAGCTCCAGGATGAGGTCTCTGGCCAGCAGCTCACCAAGGGCTCG 1537
QY 1740 AGAGCTGCTCTCTCTGGGTGTAGGCAGTGAAGCCGTGCGGGCAGAGCTGGAGAGATCAC 1799
Db 1538 GGAGCTACTCTCTGGCGTGGCGAGTGAGATCTCTCAGGCGGAGCTGCTGAAGATCAC 1597
QY 1800 AGCAGCGCCAAAACATCTGATGTCTACTCGGATCCTCAGGATCTGTTCAACCAATCCC 1859
Db 1598 CGTAGCCCCGAACATGTGATGTTCACACAGACACCTCAGGACCTGT--CAGGCCAAATCC 1655
QY 1860 TGAGCTGCAGGGGAAGCTGTGCAGCGCGGACAGCGGCTGCGGACACACAAGCCCTGGA 1919
Db 1656 AGAGCTGCAGAGGAGCTATGACGCCAGCCACCGCAGGCTGCCAGGCACAGTCACTGGA 1715
QY 1920 CACTGCTTTCATGTTGGACACCTCTGCTCAGTAGGGCCCGAGAAATTTTCTCAGATGCA 1979
Db 1716 CCTGGTCTTCTCTG-TGGATGCTCTG-CTCTGTGGGACGTGAGAACTTTGCCCAAATGCA 1773
QY 1980 GAGCTTGTGAGAGCTGTGCCCTCCAGTTTGAAGTGAACCTGACGTGACACAGGTGCG 2039
Db 1774 GAGCTTCATCAGGAATATGACCTCCGCTTGTGATGATCTCTGTGATGACAAAGTTGG 1833
QY 2040 CCTG 2043
Db 1834 CCTG 1837
RESULT 5
ABL34888
ID ABL34888 standard; cDNA; 1837 BP.
XX
AC ABL34888;
XX
XX 04-APR-2002 (first entry)
DT
XX Murine cDNA isolated from skin cells SEQ ID NO: 350.
DE
XX Human; rat; mouse; skin cell; skin wound; cancer; growth defect;
KW developmental defect; inflammatory disease; dermatological; vulnery;
KW immunomodulator; anti-inflammatory; cytostatic; neuroprotective; gene;
ss.
XX
OS Mus sp.
XX
PN WO200190357-A1.
XX
PD 29-NOV-2001.
XX
PF 24-MAY-2001; 2001WO-NZ00099.
XX

```
PR 24-MAY-2000; 2000US-206650P.
PR 25-JUL-2000; 2000US-221232P.
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX Watson JD, Strachan L, Sleeman M, Onrust R, Murison JG, Kumble KD;
XX WPI; 2002-122020/16.
XX New polynucleotides and polypeptides encoded by the polynucleotides
XX isolated from skin cells, useful for treating skin wounds, cancers,
XX growth and developmental defects, inflammatory diseases, or for
XX modulating immune responses
XX
XX Claim 1; Page 227-228; 466pp; English.
XX
XX The present invention provides the protein and coding sequences of cDNAs
XX isolated from human, murine and rat skin cell libraries. The sequences
XX can be used in the development of therapeutic agents useful in the
XX treatment of skin diseases, including skin wounds, cancer, growth
XX defects, developmental defects and inflammatory diseases. The proteins
XX have important roles in the induction of hair growth, cell proliferation
XX and cell-cell interaction, in maintaining tissue integrity, in wound
XX healing and in modulating immune responses. The present sequence is a
XX cDNA of the invention.
XX
XX Sequence 1837 BP; 370 A; 541 C; 557 G; 369 T; 0 other;
XX
XX Query Match 33.1%; Score 1117.2; DB 24; Length 1837;
XX Best Local Similarity 79.8%; Pred. No. 3.9e-267;
XX Matches 1392; Conservative 0; Mismatches 343; Indels 9; Gaps 6;
XX
XX QY 302 CGCGGCTCTCCGTTATATCAAGATGCCCTTCTCTGTCGTGGAGGCGGTCTG 361
XX DB 101 CGCGGCGTCCACAGGCTCTAACAACATGCTCCACTTCTCTACCAGCCATCTACA 160
XX
XX QY 362 TTTTCTCTGTTTCCAGATGCCCTCTCTCTCTCCAGAGTCCATGTAAACAAG 421
XX DB 161 TGTCTCTGTTTTCAGAGTGTCCCGACCATCTCTCTTCCAGGAAGTGCATGTAAACGGG 220
XX
XX QY 422 AAACCATCGGGAAGATTTACAGTCCAGCAAAATGATGTGCTCGGCTGCAGTGGACA 481
XX DB 221 AGACCATGGGAAGATCGCTGTGGCCAGCAAAATATATGTGTCAGCCGCGTGCACA 280
XX
XX QY 482 TCATGTTTCTGTAGATGGGTCTAAGCGTCCGGAAAGAGGAGCTTTGAAGTCCAAGC 541
XX DB 281 TCCTGTGTTCTGTAGATGGTCTCTACAGCATCGGGAAGGAGGAGCTTCGAGAGTCCAAGC 340
XX
XX QY 542 ACTTTGCCATCAGTCTGTGACGGTCTGGACATCAGCCCGAGAGGGTCAGAGTGGAG 601
XX DB 341 GCTTCGCCATCGCTGTGATGGCTTGACATCAGCCCTGGCAGGGTCAGAGTCCGAG 400
XX
XX QY 602 CATTCAGTTTCACTTCCACTCTCATCTGGAATTCCTCTGATTCATTTTCAACCCAAC 661
XX DB 401 CCTTGCAGTTTGTCTCACTCTCATCTGGAATTCCTCTGACTCCTTCTCAACTCGAC 460
XX
XX QY 662 AGGAAGTGAAGCAAGATCAAGAGGATGTTTTCRAAGAGGCGGCAGGAGCGGAAC 721
XX DB 461 AGGAAGTGAAGCAAGATCAAGAGGATAGTTTTCRAAGAGTGGCGCACCGAGCGGGCC 520
XX
XX QY 722 TTGCTCTGAATACCTTCTGCACAGAGGTTGCTCGAGGSCAGAAATGCTCTGTGCCCC 781
XX DB 521 TAGCCCTGAAA---CGCCTGAGCAGAGGTTTCCCGGAGGAGCAAAATGCTCTGTGCCCC 577
XX
XX QY 782 AGATCTCTCATCATCGTCACTGATGGGAAGTCCCGAGGGGATGTGGCACTGCCATCCAAGC 841
XX DB 578 AGATCTCTCATCATCGTCACTGATGGGAAGTCCCGAGGGGCGGCTCTCCCGGCTAAGC 637
XX
XX QY 842 AGCTGAAGGAAGGGGTGCTACTGTGTTGCTGTGGGGGTACAGTTTCCAGGTGGGAGG 901
XX DB 638 AGCTGAAGGAAGGGGATCGTCTGTGTTGCTGTAGAGTCCGCTTTTCCAGGTGGGAGC 697
XX
XX QY 902 AGCTGATGACCTGGCCAGGAGCCTTAGAGGGGAGCAGCAGTGTGTTGGCTGAGCAGGTGG 961
XX
XX Db 698 AGCTGCTACGCTGGCCAGTGAGCCGAAGACCGGATGTGCTGTGGCTGAGCAAGTGG 757
XX
XX QY 962 AGGATGCCACCAAGCGCTCTTTCAGCACCCCTCAGCAGCTCGGCCATCTGCTCCAGCGCCA 1021
XX
XX Db 758 AGGATGCCACCAATGGCTCTCCTCAGCACCCCTCAGCAGCTCCGACTCTGCACCAGTCTG 817
XX
XX QY 1022 CGCCAGACTGACAGGTCGAGGCTCACCCCTGTGTAGCACAGGAGCGTGGAGATGGTCCGGG 1081
XX
XX Db 818 ATCCAGACTGACAGGTCGAACTCATCCCTGTGTAGCGGAGAGCGTGGAGAGACGTCAGG 877
XX
XX QY 1082 AGTTCGCTGGCAATGCCCATGCTGTGAGAGGATCGCGGCGGACCCCTGCGGTGCTGGCTG 1141
XX
XX Db 878 AGCTGCTGGCAATGCCCTTGTGTGTGAGAGGATCAAGCAAGCAGACACTGTGCTGGCTC 937
XX
XX QY 1142 CACACTGTCCCTTCTACAGCTGGAAGAGAGTGTTCCTAACCCACCCTGCCACCTGCTACA 1201
XX
XX Db 938 TGCCCTGTCCCTTCTACAGCTGGAAGAGAGTGTTCAGACACACACCCCTGCCAATGCTACA 997
XX
XX QY 1202 GGACCACTGCCCAGGCGCTGTGACTCGCAGCCCTGCCAGATGAGGAGCAGATGTGTTTC 1261
XX
XX Db 998 GAACCATCTGTCCAGGCGCTGTGACTCCAGCCCTGCCAATGAGGAGCAGTGTGCAATTC 1057
XX
XX QY 1262 CAGAAGGACTGGACGCTACCACTGCTGCTGCGGCTGGCCCTTTGGAGGGAGGCTAACT 1321
XX
XX Db 1058 CAGAAGTGTGGATAGTAGTACACTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1117
XX
XX QY 1322 GTGCCCTGAAGCTGAGCTGGAATGCAAGTGTGACCTCTCTTCTCTGCTGCTGCTGCTGCTG 1381
XX
XX Db 1118 GTGCCCTGAAGCTGAGCTGGAATGCAAGTGTGACCTCTCTTCTCTGCTGCTGCTGCTGCTG 1177
XX
XX QY 1382 CGGCGACCACTGTGAGGCGCTTCTGCGGCGCCAAAGTCTTCGTGAAGCGGTTTGTCGGG 1441
XX
XX Db 1178 CAGGCGACCACTGTGAGGCGCTTCTGCGGCGCCAAAGCTTGTGCAAGCGCTTGTGCGAG 1237
XX
XX QY 1442 CGTGTGTGAGGAGGACTCTCGGCGCCGAGTGGGTGTGCGCACATACAGCAGGAGCTGC 1501
XX
XX Db 1238 CGTGTGTGAGGAGGACTCTCGGCGCCGAGTGGGTGTGCGCACATACAGCAGGAGCTGC 1297
XX
XX QY 1502 TGTGTGCGGTG-CCTGTGGGGAGTACCAGGA-TGTGCTCTGACCTGCTGTGAGCGCTCGA 1559
XX
XX Db 1298 TGTGTGCGGTGCTGTGCGGGAGTACCAGCATGTGTGCGGAGCTGATCAGGAGCTTGA 1357
XX
XX QY 1560 TGGCATTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1619
XX
XX Db 1358 CAGCATTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1417
XX
XX QY 1620 GCTGTGCTTCTGCGGAGCGCCACAGGAGCAGGCGCCAGCTGAGAGTGTGCTGCTGCTGCTG 1679
XX
XX Db 1418 GCACGCTTGTGGAGTGTGCGAGGAGTGTGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTG 1477
XX
XX QY 1680 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1739
XX
XX Db 1478 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1537
XX
XX QY 1740 AGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1799
XX
XX Db 1538 GGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1597
XX
XX QY 1800 AGGACGCGCCAAAGCATGTGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1859
XX
XX Db 1598 CGGTAGCGCGAAGCATGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1655
XX
XX QY 1860 TGAGCTGCGAGGGAAGCTGTGAGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1919
XX
XX Db 1656 AGAGCTGCGAGGAGGAGCTATGAGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1715
XX
XX QY 1920 CCTCGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1979
XX
XX Db 1716 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1773
XX
XX QY 1980 GAGCTTGTGAGAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2039
XX
```


Db 1774 GAGCTTCATCAGGAAATGCACCCCTCCGTTTGTATGTGAATCCTGTGATGTGACACAAGTTGG 1833
Qy 2040 CCTG 2043
|||||
Db 1834 CCTG 1837
RESULT 6
AAS81529/c
ID AAS81529 standard; cDNA; 406 BP.
XX AC AAS81529;
XX AC AAS81529;
XX AC AAS81529;
DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #17333.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0849167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX P-PSDB; ABG17342.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX PS Claim 1; SEQ ID NO 17333; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 406 BP; 80 A; 112 C; 107 G; 107 T; 0 other;
Query Match 10.1%; Score 340; DB 23; Length 406;
Best Local Similarity 90.7%; Pred. No. 2.1e-74;
Matches 390; Conservative 0; Mismatches 0; Indels 40; Gaps 1;

Qy 2482 AGCCCGTCATGAATGAGGGCAGCTGCTCTCTGCAGAAATGGAGCTACCGCTGCAAGTGT 2541
Db 406 AGCCCGTCATGAATGAGGGCAGCTGCTCTCTGCAGAAATGGAGCTACCGCTGCAAGTGT 347
Qy 2542 CGGGATGGCTGGGAGGGCCCCCACTGCGAGAACCGTGAAGTGGAGCTCTTGCCTCTGTATGT 2601
Db 346 CGGGATGGCTGGGAGGGCCCCCACTGCGAGAACCGTGAAGTGGAGCTCTTGCCTCTGTATGT 313
Qy 2602 GTGAGCCAGGATGATGATTCTTTGAGACGCCCTGTAGGCACATGGCTCCCGTGCAGGAGGGC 2661
Db 312 -----GATTCTTGAGACGCCCTGTAGGCACATGGCTCCCGTGCAGGAGGGC 267
Qy 2662 AGCAGCGGTACCCCTCCAGCAACTACAGAGAGCGCTGGGCACTGAAATGTCCTTACC 2721
Db 266 AGCAGCGGTACCCCTCCAGCAACTACAGAGAGCGCTGGGCACTGAAATGTCCTTACC 207
Qy 2722 TTCTGGAAATGCTGTCGCCCGCAGGTCCTTAGAATGCTCTGCTCCCGCTGCGCAGGACCA 2781
Db 206 TTCTGGAAATGCTGTCGCCCGCAGGTCCTTAGAATGCTCTGCTCCCGCTGCGCAGGACCA 147
Qy 2782 CTATTCTCACTGAGGAGGAGGATGTCCCAACTGCAGCCATGCTGCTTAGAGACAAGAAA 2841
Db 146 CTATTCTCACTGAGGAGGAGGATGTCCCAACTGCAGCCATGCTGCTTAGAGACAAGAAA 87
Qy 2842 GCAGCTGATGTCAACCCACAAACGATGTTGTTGAAAGCTTTTGTAGTGTAAATATACC 2901
Db 86 GCAGCTGATGTCAACCCACAAACGATGTTGTTGAAAGCTTTTGTAGTGTAAATATACC 27
Qy 2902 ACTTTCGTGA 2911
Db 26 ACTTTCGTGA 17
RESULT 7
ABQ16108/c
ID ABQ16108 standard; DNA; 651 BP.
XX AC ABQ16108;
XX AC ABQ16108;
DT 12-JUL-2002 (first entry)
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 2699.
XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX OS Homo sapiens.
XX WO200218632-A2.
XX 07-MAR-2002.
XX 01-SEP-2001; 2001WO-EP10074.
XX 01-SEP-2000; 2000DE-1043826.
XX 05-SEP-2000; 2000DE-1044543.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
PI WPI; 2002-371829/40.
XX DR Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

XX Sequence 651 BP; 112 A; 68 C; 227 G; 244 T; 0 other;

Query Match 6.0%; Score 201; DB 24; Length 651;

Best Local Similarity 76.6%; Pred. No. 8.1e-40;

Matches 246; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 23 CTCGGAGGCTGGGTGACCGCGTAGAAGTGAAGTACTTTTATTGTCAGACCTGGGCGG 82

Db 651 CTCGAAACTAAATAACCGCGTAAAAATAAAATACTTTTATTACAAACCTAAACCG 592

QY 83 ATGCGCGCTTTAAACACGCGAGGGGCTCTATGCACCTCCCTGGCGTAGTCTCCGACC 142

Db 591 ATACCGCTTTTAAACACGCGAAGAACTCTATACACCTCCCTAACGATAATCTCCGACC 532

QY 143 TCAGCGGGTGGGTGCTGCCGCCCTCTCCAGGAGAGACAAACAGGTGTCCACGTGGC 202

Db 531 TCACCGAATCGATCGTACCGCCTCTCCAAAAACAAACAAATATCCACGTAC 472

QY 203 AGCGGGGCGCGGGCGCCCTCTGTGATCCCGTAGCGCCCTCGCGAGCGCGCGCC 262

Db 471 AACCGCGCGCGAAGCGCCCTCTATATCCGTAACGCCCTAACCGAAGCGCGCC 412

QY 263 GGGTGTGTGAGTAGAGCGCGCGCGAGCGGTGTGTCGCGCTCTCTTCGGTTATA 322

Db 411 GAATCTATAATAAACAACCGCGCGAAGCGTAATCGCGCTCTCTTCGGATAA 352

QY 323 TCACATGCCCCCTTTCTGT 343

Db 351 TCCCAACCCCGAATTAACCTCT 331

RESULT 8

ABQ16109

ID ABQ16109 standard; DNA; 651 BP.

XX AC

XX ABQ16109;

XX DB

XX 12-JUL-2002 (first entry)

XX DE

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 2700.

XX KW

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

XX KW

XX drug; side effect; cancer; central nervous system; cardiovascular;

XX KW

XX gastrointestinal; respiratory system; single nucleotide polymorphism;

XX KW

XX SNP; cell differentiation; ds.

XX OS

XX Homo sapiens.

XX WO200218632-A2.

XX PN

XX 07-MAR-2002.

XX PD

XX 01-SEP-2001; 2001WO-EP10074.

XX PF

XX

XX

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX (EPIG-) EPIGENOMICS AG.

PA Olek A, Piepenbrock C, Berlin K, Guetig D;

PI WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful

XX for diagnosis and prognosis, comprises selective hybridization of

XX amplicons from chemically treated DNA

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of

XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a

XX genomic sample of DNA. The sample is treated chemically to convert

XX cytosine (C) but not methylated C, to uracil, then part of the genomic

XX DNA that contains the target C is amplified to form a labeled amplicon.

XX The amplicon is hybridised to two classes, each with at least one

XX member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers

XX and the degree of hybridisation to both classes is determined from the

XX label on the amplicon. From the ratio of labels hybridised to the two

XX classes of oligomers, the degree of methylation is calculated. The method

XX is used: (i) for diagnosis and/or prognosis of side effects of

XX therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders

XX of the central nervous, cardiovascular, gastrointestinal and respiratory

XX systems etc., particularly by detecting mutations or single nucleotide

XX polymorphisms (SNP's); and (ii) for differentiation of cell or tissue

XX types and for investigating cell differentiation. The method allows the

XX methylation status of many C residues to be determined simultaneously.

XX ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the

XX method for determining the degree of cytosine methylation described in

XX the disclosure of the invention.

XX Sequence 651 BP; 244 A; 227 C; 68 G; 112 T; 0 other;

Query Match 6.0%; Score 201; DB 24; Length 651;

Best Local Similarity 76.6%; Pred. No. 8.1e-40;

Matches 246; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 23 CTCGGAGGCTGGGTGACCGCGTAGAAGTGAAGTACTTTTATTGTCAGACCTGGGCGG 82

Db 1 CTCGAAACTAAATAACCGCGTAAAAATAAAATACTTTTATTACAAACCTAAACCG 60

QY 83 ATGCGCGCTTTAAACACGCGAGGGGCTCTATGCACCTCCCTGGCGTAGTCTCCGACC 142

Db 61 ATACCGCTTTAAACACGCGAAGAACTCTATACACCTCCCTAACGATAATCTCCGACC 120

QY 143 TCAGCGGGTGGGTGCTGCCGCCCTCTCCAGGAGAGACAAACAGGTGTCCACGTGGC 202

Db 121 TCAACCGAATCGAATCGTACCGCCTCTCCAAAAACAAACAAATATCCACGTAC 180

QY 203 AGCGGGGCGCGGGCGCCCTCTGTGATCCCGTAGCGCCCTCGCGAGCGCGCGCC 262

Db 181 AACCGCGCGCGAAGCGCCCTCTATATCCGTAACGCCCTAACCGAAGCGCGCC 240

QY 263 GGGTGTGTGAGTAGAGCGCGCGCGAGCGGTGTGTCGCGCTCTCTTCGGTTATA 322

Db 241 GAATCTATAATAAACAACCGCGCGAAGCGTAATCGCGCTCTCTTCGGATAA 300

QY 323 TCACATGCCCCCTTTCTGT 343

Db 301 TCCCAACCCCGAATTAACCTCT 321

RESULT 9

ABQ16106

ID ABQ16106 standard; DNA; 651 BP.

XX AC

XX ABQ16106;

XX

Best Local Similarity 58.8%; Pred. No. 7.4e-34;
Matches 243; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 24 TCGAGGCTGGGTACCCCGTAGAAGTGAATCTTTTATTTCAGACTGGGCCGA 83
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 650 TCGAGGCTGGGTGATTTCGCGTAGAAGTGAAGTAATTTTTATTGTAGATTTCGGTGCA 591

QY 84 TGCCGGCTTTAAAAACGCAGGGGCTCATGCACTCCTCGCGGTAAGTTCCCTCCGACC 143
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 590 TGTCTGTTTTAAAAAACGCAGGGGTTTTATGTAATTTTTTGGCGGTAGTTTTTCGATT 531

QY 144 CAGCGGCTGGGTGCGTGC CGCCCTCTCCCAGGAGACAACAAGGTGTCACCGTGCCA 203
|| | ||||| ||||| || | ||||| ||||| ||||| ||||| |||||
Db 530 TAGTCGGGTCGGGTGCTGCTTTTATTTAGGAGAGATAAATAGGTGTTTACGTGGTA 471

QY 204 GC CGCGCCCCGGCGGCCCTCCTGTGATCCCGTAGCGCCCTGGCCGAGCCGCGCCCG 263
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 470 GTCGGTTTCGGGCGTTTTTTTGTGATTTCTGAGCGTTTTTGGTTCGAGTCGCGTTTCG 411

QY 264 GGTCGTGTAGTAGAGCGCCCGCACCGAGCGGTGTCGCGGCTCTCCTTCCTGTTATAT 323
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 410 GGTTCGTGTAGTAGAGTCGTTTCGGGTATCGAGCGTGGTCGCTGTTTTTTTCGGTGA 351

QY 324 CAACATSCCCCTTCCTGTGTCGGAGCGCGTCTGTGTTTCCTGTTTCCCA 376
|| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 350 TTGAGTTCAGTTCGTTTTGGCGTTCGGTAATTCGTAGTTCGGTTGTCGA 298

RESULT 11
AAS91851
ID AAS91851 standard; cDNA; 1491 BP.
XX AC AAS91851;
XX XX
X 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #27655.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX PA {HYSE-} HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR P-PSDB; ABG27664.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX PS Claim 1; SEQ ID NO 27655; 103pp; English.

(II).. (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (ii) and its binding partners are useful in medical imaging of sites expressing (ii). (I) and (ii) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 1491 BP; 307 A; 417 C; 474 G; 293 T; 0 other;

Query Match 2.9%; Score 98.8; DB 23; Length 1491;
Best Local Similarity 50.8%; Pred. No. 2.6e-14;
Matches 266; Conservative 0; Mismatches 252; Indels 6; Gaps 1;

QY 422 AAACCATCGGGAAGATTTCAGCTGCCAGCAAAATGATGTGGTCTCGGCTGCAGTAGTGACA 481
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 767 ACAGCAGCGCAAGACTCGCAATCTCTCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 826
QY 482 TCATGTTTCTGTAGATGGGTCTAACAGCGTCGGGAAGGAGCTTTGAAAGGTCCCAAGC 541
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 827 TGGTCTTCCTCANTTGACGGATCCAAGATGTGAGCCAGAGAAGTAATTTGAGCTGGTGAAGA 886
QY 542 ACTTTGCCATCACAGTCTGTGACGGTCTGGACATCAGCCCCGAGAGGGTCAAGTGGGAG 601
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 887 AGTTCATCAGTCAGATCGTGGATACGCTGGACGTGCAGACAAGCTGGCCACGGTGGGGC 946
QY 602 CATTCAGTTCAGTTCACCTCCTCATCTGCAATTCCTTGGATTCATTTCAACCCAAC 661
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 947 TGGTGCAGTACTCAAGCTCTGTGCGCCAGAGTGTCCCCTGGTGGCTCCACACCAGA 1006
QY 662 AGAAGTGAAGGCAAGAATAACAAGAGATGTTTTCAAGGAGGGCGCACGAGACGGAAC 721
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1007 AGGACATCAAGCGCGGTGTCCGGAATATGCTTACATGGAGAAGGGCACAATGACTGGGG 1066
QY 722 TTGCTGTGAATACCTTCTCCAGAGGGTGTGCTGGAGGACAAATGCT-----TCTG 775
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1067 CTGCTCTCAAGTACCTCAITTGACAATTCCTTCACTGTGTCCAGTGGGGCTAGGCCCGGG 1126
QY 776 TGCCCCAGTCCCTCATCTCATCTGATGGGAAGTCCCAGGGGATGTGGCATGCCAT 835
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1127 CCAGAAGTGGCATGTCTTCACTGATGCCGAGCCAGACTACATTAATGATGCTG 1186
QY 836 CCAAGCAGCTGAAGGAAGGGGTGTCACTGTGTGTGCTGGGGGTGAGGTTCACAGGT 895
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1187 CCAAGAGGCAAGAGACTCGGCTTTAAGATGTTGCTGTGGTGTGGGCAATGCCGTGG 1246
QY 896 GGAGGAGCTGCATGCAGCTGGCAGGAGCCTAGAGGCGAGCAC 939
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1247 AGGATGAGCTGAGGGAATAGCCTCAGAGCCTGTGGCAGAGCAC 1290

RESULT 12
AAS94773
ID AAS94773 standard; DNA; 4151 BP.
XX
XX AAS94773;
XX AC
XX AT
XX DT
XX DE
XX DE
XX DE
KW Human; foam cell differentiation; atherosclerosis; cerebral stroke;
KW cardiovascular disorder; coronary artery disease; gene therapy; ds.
XX OS
XX OS Homo sapiens.

PN W0200177389-A2.
XX 18-OCT-2001.
XX 04-APR-2001; 2001WO-US11128.
XX 05-APR-2000; 2000US-195106P.
XX (INCY-) INCYTE GENOMICS INC.

XX Shiffman D, Somogyi R, Lawn R, Seilhamer JJ, Porter GJ, Mikita T;
PI Tai J;
XX WPI: 2002-010925/01.

XX Composition useful for diagnosis of conditions, disorders or diseases
PT associated with atherosclerosis, comprises several polynucleotides that
PT are differentially expressed in foam cell development.

PS Claim 1; Page 84-85; 315pp; English.

XX The present invention relates to the isolation of human polynucleotide
CC sequences that are differentially expressed during foam cell
CC differentiation. The polynucleotide sequences of the invention or a
CC composition comprising these polynucleotides are useful as a high
CC throughput method for detecting altered expression of one or more
CC polynucleotides in a sample. The polynucleotides can be used in the
CC diagnosis of disorders associated with foam cell development such as
CC atherosclerosis, cerebral stroke, and cardiovascular disorders such as
CC coronary artery disease. The polynucleotide sequences can also be used
CC as PCR primers and probes. The polynucleotides of the invention are also
CC useful in gene therapy. AAS94746-AAS95021 represent the human
CC polynucleotide sequences of the invention which are differentially
CC expressed during foam cell differentiation.

XX Sequence 4151 BP; 889 A; 1089 C; 1163 G; 928 T; 82 other;

Query Match 2.9%; Score 97.8; DB 24; Length 4151;
Best Local Similarity 50.6%; Pred. No. 7e-14;
Matches 265; Conservative 0; Mismatches 253; Indels 6; Gaps 1;

QY 422 AACCATCGGAGATTCAGCTCCAGCAAAATGATGTGCTCGGCTCCAGTGACA 481
DB 1128 ACAGCGACGACGACCTGCAATGCTGTCAGTGTGTGGTGGCAGCTCGGCCACTGACC 1187
QY 482 TCATGTTTCTGTAGATGGGTCTAACAGCGTGGGAAAGGAGGCTTTGAAAGGTCCAAGC 541
DB 1188 TGGTCTTCTTATGACGGATCCAGAGTGTGAGCCAGAGAACCTTGAGCTGGTGAGA 1247
QY 542 ACTTTGCCATCACAGTCTGTGACGGTCTGGACATCAGCCCCGAGAGGGTCCAGAGTGGGAG 601
DB 1248 AGTTTCATCAGTCAGATCGTGATACGCTGGAGTGTGACAGCAAGCTGGCCAGGTGGGC 1307
QY 602 CATTCAGTTTCACTTCACCTCATCTGGAATTCCTTGGATTTCATTTCACCCCAAC 661
DB 1308 TGGTGCATCTCAAGCTCTGTGCGCCAGGAGTTCCCTCCCTGGCTTCCACACCAAGA 1367
QY 662 AGGAAGTGAAGGCAAGAAATCAAGAGGATGTTTTCAAAGGAGGCGCCAGGACGGAAC 721
DB 1368 AGGACATCAAGCGGCTGTGCGGATATGCTTACATGGAGAGGGGCAATGACCCGGG 1427
QY 722 TTGCTCTGAATACCTTCTCCACAGAGGGTTGCCCTGGAGGAGAAATGCT-----TCTG 775
DB 1428 CTGCTCTCAAGTACCTCATTCAGCAATTCCTTCACTGTGTCCAGTGGNGCTAGGCCCGGG 1487
QY 776 TGCCCCAGATCCCATCATCTGCTACTGATGGAGTCCAGGGGATGTTGCCACTGCCAT 835
DB 1488 CCCAGAAGGGTGGCATTTGCTTCACTATGCGCGGAGGACTACATTAATGATGCTG 1547
QY 836 CCAAGCAGCTGAAGGAAAGGGTGTCACTGTGTTTGTCTGTGGGGTCCAGGTTTCCCAAGT 895
DB 1548 CCAGAAGGCCAAGAGCTCGGCTTTAAGATGTTTGTGTGGGTGTGGCAATGCCCTGG 1607

QY 896 GGGAGGAGTGCATGTCACCTGCCAGCGAGCTAGAGGCGACGAC 939
DB 1608 AGGATGAGTGCAGGGAATAGCCTCAGAGCTGTGGCAGAGCAC 1651

RESULT 13

AA000664
ID AAX00664 standard; DNA; 1668 BP.

XX AAX00664;

XX 25-MAR-1999 (first entry)

XX Human secreted protein gene 54 clone HE8EM69.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX Homo sapiens.

XX W09842738-A1.

XX 01-OCT-1998.

XX 19-MAR-1998; 98WO-US05311.

XX 30-MAY-1997; 97US-0050937.

XX 21-MAR-1997; 97US-0041276.

XX 21-MAR-1997; 97US-0041277.

XX 21-MAR-1997; 97US-0041281.

XX 21-MAR-1997; 97US-0042344.

XX 30-MAY-1997; 97US-0048069.

XX 30-MAY-1997; 97US-0048094.

XX 30-MAY-1997; 97US-0048095.

XX 30-MAY-1997; 97US-0048096.

XX 30-MAY-1997; 97US-0048099.

XX 30-MAY-1997; 97US-0048131.

XX 30-MAY-1997; 97US-0048135.

XX 30-MAY-1997; 97US-0048154.

XX 30-MAY-1997; 97US-0048160.

XX 30-MAY-1997; 97US-0048186.

XX 30-MAY-1997; 97US-0048187.

XX 30-MAY-1997; 97US-0048188.

XX 30-MAY-1997; 97US-0048350.

XX 30-MAY-1997; 97US-0048351.

XX 30-MAY-1997; 97US-0048352.

XX 30-MAY-1997; 97US-0048355.

XX 05-AUG-1997; 97US-0054804.

(HUMA-) HUMAN GENOME SCI INC.

Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA;

Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS;

Rosen CA, Ruben SM, Shi Y, Young P;

WPI: 1999-070066/06.

P-PSDB; AAW67860.

XX New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders

PS Claim 1; Page 215-216; 385pp; English.

XX This sequence represents a nucleic acid molecule which encodes a
CC secreted human protein. The gene number, and the clone it is derived
CC from, are detailed in the descriptor line. The gene can be used to

Db 414 AGCGCAGGTCACGCCCGCGTCCCTTAACGTGCTCCTGTTGGCTGCTTTCACGG 473
QY 2249 GCGGAGAGCGGCGGAGGATGCAGCGCTTCTCTGCCACAGAGCTGAGGAACAATGGCATCT 2308
Db 474 ATGCCCTCCAGGATGACATCTCGTGTGGGCGAGCGCGGCAAGAGGAGGATCG 533
QY 2309 CTGCTTTGGTGGGCGTGGGCGCTCTCCTTAAGTGAAGGCTCTGCGAGGCTTTC 2363
Db 534 TCATGTACGCGCTGGGCGTGGGCAAGCGGTGGAGGCGGAGCTGCGCGAGATCOC 588

RESULT 15
AAH17154
ID AAH17154 standard; cDNA; 1167 BP.
AC AAH17154;
XX
XX 26-JUN-2001 (first entry)
XX Human cDNA sequence SEQ ID NO:16503.
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
OS Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 16503; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.
XX
SQ Sequence 1167 BP; 203 A; 337 C; 424 G; 203 T; 0 other;
Query Match 2.2%; Score 74.2; DB 22; Length 1167;
Best Local Similarity 49.1%; Pred. No. 3.1e-08;
Matches 233; Conservative 0; Mismatches 233; Indels 9; Gaps 1;
QY 1898 GGTGCCGACACAAAGCCCTGACCTCTCTTTCATGTTGGACACCTCTGCTCAGTAGGC 1957
Db 114 GGTGCCGGAAGCCACGCTGACCTTCTTCTGCTGTTGATGGCTCCAAGAGCGTGGCT 173
QY 1958 CCGAGAAATTTTGTCTAGATCAGAGCTTTGTGAGAAGCTGTGCGCTCCAGTTTGAAGTGA 2017
Db 174 CACAAAACCTTCGAGCTAGTGAAGCGCTTCGTGAACACAGATTGTGGACTTCTCTAGATGT 233
QY 2018 ACCTGACGTGACACAGTCTGGCTGTGTATGTCACGACGAGTGCAGTCTGCTTCG 2077
Db 234 CCCCCGAGGACGCGGCTGGCTGTCAGTTCGAGCCGCTGCGCACCGAGTTCC 293
QY 2078 GGCTGGACACCAACCCACCCGGCTGCGATGCTGCGGGCCATTAGCCAGGCCCCCTACC 2137
Db 294 CTCTGGTCTCTAGCGCACCCAGCCAGCTGAAGCAGCGGCTCTGCGCTGGAGTACA 353
QY 2138 TAGGTGGGTGGGCTCAGCCGCGCACCCCTGCTGTCACATCTATGACAAAGTGTATGACCG 2197
Db 354 TGGAAACGCGGCACCATGACAGGGCTGGCTTGGCGCACATGTTGGAGCACAGCTTCTCCG 413
QY 2198 TCCAGAGGGTGGCGGCTGCTGCTCCCAA-----AGCTGTGGTGGTCTCACAG 2248
Db 414 AGCGCAGGTGTCACGCCCGCTTAACTGTCCTGTTGGCTTGTCTTTCACGG 473
QY 2249 GCGGAGAGCGGACAGAGGATGCAGCCCTTCTCTGCCAAGAGCTGAGGAACAATGGCATCT 2308
Db 474 ATGCCCTCCAGGATGACATCTCGTGTGGGCGAGCGCGCCCAAGGAGGAGGATCG 533
QY 2309 CTGCTTTGGTGGGCGTGGGCGCTCTCCTTAAGTGAAGGCTCTGCGGAGGCTTTC 2363
Db 534 TCATGTACGCGCTGGGCGTGGGCAAGCGGTGGAGGCGGAGCTGCGCGAGATCOC 588

Search completed: May 5, 2003, 02:27:15
Job time : 493 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 02:27:24 ; Search time 248 Seconds
(without alignments)
16080.529 Million cell updates/sec

Title: us-09-930-020a-1

Perfect score: 3375

Sequence: 1 gacagtgttcgcgctgcac.....tgtccaccttgaagtgcttc 3375

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3375	100.0	3375	9	US-09-930-020a-1
2	1117.2	33.1	1837	9	US-10-152-661-350
3	1117.2	33.1	1837	9	US-09-866-050A-350
4	74.6	2.2	1668	9	US-09-984-245-64
5	74.6	2.2	1668	9	US-09-966-262-64
6	74.6	2.2	1668	9	US-09-983-966-64
7	74.6	2.2	1668	9	US-10-143-090-64
8	73.6	2.2	3447	9	US-10-000-512-9
9	73.6	2.2	3449	9	US-09-905-291A-33
10	73.6	2.2	3449	9	US-09-902-853-33
11	73.6	2.2	3449	9	US-09-907-824-33
12	73.6	2.2	3449	9	US-09-907-841-33
13	73.6	2.2	3449	9	US-09-904-011-33
14	73.6	2.2	3449	9	US-10-028-072-293
15	73.6	2.2	3449	9	US-09-906-742-33
16	73.6	2.2	3449	9	US-10-121-049-293
17	73.6	2.2	3449	9	US-10-123-904-293
18	73.6	2.2	3449	9	US-10-140-470-293
19	73.6	2.2	3449	9	US-09-906-838-33

20	73.6	2.2	3449	9	US-09-907-613-33	Sequence 33, Appl
21	73.6	2.2	3449	9	US-09-907-942-33	Sequence 33, Appl
22	73.6	2.2	3449	9	US-10-175-746-293	Sequence 293, App
23	73.6	2.2	3449	9	US-10-176-918-293	Sequence 293, App
24	73.6	2.2	3449	9	US-10-176-921-293	Sequence 293, App
25	73.6	2.2	3449	9	US-10-137-865-293	Sequence 293, App
26	73.6	2.2	3449	9	US-10-140-474-293	Sequence 293, App
27	73.6	2.2	3449	9	US-09-904-820-33	Sequence 33, Appl
28	73.6	2.2	3449	9	US-09-904-859-33	Sequence 33, Appl
29	73.6	2.2	3449	9	US-09-909-204-33	Sequence 33, Appl
30	73.6	2.2	3449	9	US-10-142-431-293	Sequence 293, App
31	73.6	2.2	3449	9	US-10-143-114-293	Sequence 293, App
32	73.6	2.2	3449	9	US-09-904-786-33	Sequence 33, Appl
33	73.6	2.2	3449	9	US-09-906-646-33	Sequence 33, Appl
34	73.6	2.2	3449	9	US-09-906-700-33	Sequence 293, App
35	73.6	2.2	3449	9	US-10-140-002-293	Sequence 33, Appl
36	73.6	2.2	3449	9	US-09-902-903-33	Sequence 33, Appl
37	73.6	2.2	3449	9	US-09-903-749A-33	Sequence 33, Appl
38	73.6	2.2	3449	9	US-09-903-786-33	Sequence 33, Appl
39	73.6	2.2	3449	9	US-10-142-419-293	Sequence 293, App
40	73.6	2.2	3449	9	US-09-902-736-33	Sequence 33, Appl
41	73.6	2.2	3449	9	US-09-904-119-33	Sequence 33, Appl
42	73.6	2.2	3449	9	US-09-904-956-33	Sequence 33, Appl
43	73.6	2.2	3449	9	US-09-907-794-33	Sequence 33, Appl
44	73.6	2.2	3449	9	US-10-123-262-293	Sequence 293, App
45	73.6	2.2	3449	9	US-10-142-423-293	Sequence 293, App

ALIGNMENTS

RESULT 1

US-09-930-020a-1

; Sequence 1, Application US/09930020A

; Publication No. US20030077568A1

; GENERAL INFORMATION:

; APPLICANT: Glsh, Kurt C.

; APPLICANT: Mack, David H.

; APPLICANT: Wilson, Keith E.

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Colorectal Cancer, Compositions

; TITLE OF INVENTION: and Methods of Screening for Colorectal Cancer

; TITLE OF INVENTION: Modulators

; FILE REFERENCE: 018501-00310005

; CURRENT APPLICATION NUMBER: US/09/930, 020A

; CURRENT FILING DATE: 2001-08-14

; PRIOR APPLICATION NUMBER: US 09/663,733

; PRIOR FILING DATE: 2000-09-15

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 3375

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: CBF9

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (328)..(2751)

; OTHER INFORMATION: CBF9

US-09-930-020a-1

Query Match 100.0%; Score 3375; DB 9; Length 3375;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACAGTTCGCGGCTGCACCGCTCGGAGCTGGGTGACCCGCTAGCACTGAAGTACTT 60

Db 1 GACAGTTCGCGGCTGCACCGCTCGGAGCTGGGTGACCCGCTAGCACTGAAGTACTT 60

QY 61 TTTTATTTGACAGACTGGGCGGATGCCGCTTTAAAAAACGAGGGGCTCTATGCACCTC 120

Db 61 TTTTATTTGACAGACTGGGCGGATGCCGCTTTAAAAAACGAGGGGCTCTATGCACCTC 120

QY 121 CCTGGGGTAGTTCCTCCGACCTCAGCCGGTCCGGTCTGGCCGCTCTCCAGAGAG 180
DB 121 CCTGGCGGTAGTTCCTCCGACCTCAGCCGGTCCGGTCTGGCCGCTCTCCAGAGAG 180
QY 181 ACAAAAGAGTGTCCACAGTGGCAGCGCGCCCGCGGCCCTCTGTGATCCCTAGCG 240
DB 181 ACAAAAGAGTGTCCACAGTGGCAGCGCGCCCGCGGCCCTCTGTGATCCCTAGCG 240
QY 241 CCCCTTGGCCGAGCGCGCCCGGGTCTGTGATGAGCGCGCCCGGCCACCGAGCGCTGG 300
DB 241 CCCCTTGGCCGAGCGCGCCCGGGTCTGTGATGAGCGCGCGCCCGGCCACCGAGCGCTGG 300
QY 301 TGGCGCTCTCTCTCGTTATATCAACATGCCCCCTTTCTGTGTGGAGCCGCTGT 360
DB 301 TGGCGCTCTCTCTCGTTATATCAACATGCCCCCTTTCTGTGTGGAGCCGCTGT 360
QY 361 GTTTTCTCTTTTCCAGAGTCCCGCATCTCTCCCTCTCCAGGAAGTCCATGAACAAA 420
DB 361 GTTTTCTCTTTTCCAGAGTCCCGCATCTCTCCCTCTCCAGGAAGTCCATGAACAAA 420
QY 421 GAAACCATCGGAAGATTTCAGCTGCGCAGCAAAATGATGTGCTCGGCTCGAGTGGAC 480
DB 421 GAAACCATCGGAAGATTTCAGCTGCCAGCAAAATGATGTGCTCGGCTCGAGTGGAC 480
QY 481 ATCATGTTTCTGTAGATGGGTCTAACAGCGTGGGAAAGGAGCTTTGAAAGTCCAAAG 540
DB 481 ATCATGTTTCTGTAGATGGGTCTAACAGCGTGGGAAAGGAGCTTTGAAAGTCCAAAG 540
QY 541 CACTTTGCCATCACAGTCTGTGACCGTCTGGACATCAGCCCCGAGAGGTCAGAGTGGGA 600
DB 541 CACTTTGCCATCACAGTCTGTGACCGTCTGGACATCAGCCCCGAGAGGTCAGAGTGGGA 600
QY 601 GCATTCCAGTTCCAGTCCACTCTCATCTGGAATCCCTTGATTCATTTTCAACCCAA 660
DB 601 GCATTCCAGTTCCAGTCCACTCTCATCTGGAATCCCTTGATTCATTTTCAACCCAA 660
QY 661 CAGGAAGTGAAGCAAGAAATCAAGAGGATGGTTTCAAAGAGGCGCACGGAGCGGAA 720
DB 661 CAGGAAGTGAAGCAAGAAATCAAGAGGATGGTTTCAAAGAGGCGCACGGAGCGGAA 720
QY 721 CTTGCTCTCAATACCTCTGCACAGAGGTTGCCCTGGAGGAGAAATCTCTGTGCCC 780
DB 721 CTTGCTCTCAATACCTCTGCACAGAGGTTGCCCTGGAGGAGAAATCTCTGTGCCC 780
QY 781 CAGATCCTCATCATGCTCACTGATGGGAAGTCCCAGGGGATGTGGCTATGCCATCCAAAG 840
DB 781 CAGATCCTCATCATGCTCACTGATGGGAAGTCCCAGGGGATGTGGCTATGCCATCCAAAG 840
QY 841 CAGCTCAAGGAAGGGTCTCACTGTGTTGCTGTGGGGTCCAGTTCAGAGTGGGAG 900
DB 841 CAGCTCAAGGAAGGGTCTCACTGTGTTGCTGTGGGGTCCAGTTCAGAGTGGGAG 900
QY 901 GAGCTCATCTGATGCCAGCGCTAGAGGGCAGCAGTGTCTGTGGCTGAGCAGGTG 960
DB 901 GAGCTCATCTGATGCCAGCGCTAGAGGGCAGCAGTGTCTGTGGCTGAGCAGGTG 960
QY 961 GAGGATGCCAACAGCGCTCTTACAGACCCCTCAGCAGCTCGGCCCATCTGCTCCAGCGCC 1020
DB 961 GAGGATGCCAACAGCGCTCTTACAGACCCCTCAGCAGCTCGGCCCATCTGCTCCAGCGCC 1020
QY 1021 ACGCCAGACTGCAGGGTCCAGGCTCACCCCTGTGACACAGCAGCAGTGTGCTCCG 1080
DB 1021 ACGCCAGACTGCAGGGTCCAGGCTCACCCCTGTGACACAGCAGCAGTGTGCTCCG 1080
QY 1081 GAGTTCCGTGGCAATGCCCATCTGAGAGGATCGCGCGGACCCCTTCCGCTGTGGCT 1140
DB 1081 GAGTTCCGTGGCAATGCCCATCTGAGAGGATCGCGCGGACCCCTTCCGCTGTGGCT 1140
QY 1141 GCACATGTCCTCTTACAGCTGGAAGAGAGTGTTCCTTAACCCACCCCTGCCACCTGCTAC 1200
DB 1141 GCACATGTCCTCTTACAGCTGGAAGAGAGTGTTCCTTAACCCACCCCTGCCACCTGCTAC 1200

QY 1201 AGGACCCTGCCAGGCCCTGTGACTCGCAGCCCTGCCAGAAATGGAGGACATGTGTT 1260
DB 1201 AGGACCCTGCCAGGCCCTGTGACTCGCAGCCCTGCCAGAAATGGAGGACATGTGTT 1260
QY 1261 CCAGAAAGACTGACGGCTACCAAGTGCCTCTCCCGCTGGCCTTTGGAGGGAGGCTAAC 1320
DB 1261 CCAGAAAGACTGACGGCTACCAAGTGCCTCTCCCGCTGGCCTTTGGAGGGAGGCTAAC 1320
QY 1321 TGTGCCCTGAAGCTGAGCCTGGAATGCAGGGTGCACCTCTCTCTCTGCTGGACAGCTCT 1380
DB 1321 TGTGCCCTGAAGCTGAGCCTGGAATGCAGGGTGCACCTCTCTCTCTGCTGGACAGCTCT 1380
QY 1381 GCGGGACCACTCTGAGCGGCTTCTCGGGGCCAAAGTCTTCGTTGAAGCGGTTTGTGCGG 1440
DB 1381 GCGGGACCACTCTGAGCGGCTTCTCGGGGCCAAAGTCTTCGTTGAAGCGGTTTGTGCGG 1440
QY 1441 GCCTGTCTGAGGAGACTCTCGGGCCGAGTGGGTGTGGCCACATACACAGAGGAGCTG 1500
DB 1441 GCCTGTCTGAGGAGACTCTCGGGCCGAGTGGGTGTGGCCACATACACAGAGGAGCTG 1500
QY 1501 CTGTGGCGGTGCCCTGTGGGGAGTACCAAGATGTGCCCTGACCTGGTCTGGAGCCTCGAT 1560
DB 1501 CTGTGGCGGTGCCCTGTGGGGAGTACCAAGATGTGCCCTGACCTGGTCTGGAGCCTCGAT 1560
QY 1561 GGCATTCCCTTCGCTGGTGGCCGCCACCTGACGGGAGTGCCTTGGCGCAGGCGGACAG 1620
DB 1561 GGCATTCCCTTCGCTGGTGGCCGCCACCTGACGGGAGTGCCTTGGCGCAGGCGGACAG 1620
QY 1621 CTTGCTCTGGGAGCGCCACAGGACAGGCCAGGCCGACGCTAGAGTGTGTGTTTG 1680
DB 1621 CTTGCTCTGGGAGCGCCACAGGACAGGCCAGGCCGACGCTAGAGTGTGTGTTTG 1680
QY 1681 CTCCTGAGTCACTCCGAGGATGAGGTTGGGGGCCAGCGCTACGCAAGGGCGCA 1740
DB 1681 CTCCTGAGTCACTCCGAGGATGAGGTTGGGGGCCAGCGCTACGCAAGGGCGCA 1740
QY 1741 GAGTGTCTCTCTGGGTAGCAGTGAAGCGCGTGGCGGACAGCTGGAGGAGATCACA 1800
DB 1741 GAGTGTCTCTCTGGGTAGCAGTGAAGCGCGTGGCGGACAGCTGGAGGAGATCACA 1800
QY 1801 GGCAGCCCAAGCATGTGATGCTCTACTCGGATCTCTCAGGATCTGTTCACCAATCCCT 1860
DB 1801 GGCAGCCCAAGCATGTGATGCTCTACTCGGATCTCTCAGGATCTGTTCACCAATCCCT 1860
QY 1861 GAGCTGACGGGAAGTGTGACGCGGCGCAGCGGCTGCGGACACAAAGCCCTGGAC 1920
DB 1861 GAGCTGACGGGAAGTGTGACGCGGCGCAGCGGCTGCGGACACAAAGCCCTGGAC 1920
QY 1921 CTCGTCTTCATGTTGGACACCTCTGCCTCAGTAGGGCCGAGAAATTTGCTCAGATGCAG 1980
DB 1921 CTCGTCTTCATGTTGGACACCTCTGCCTCAGTAGGGCCGAGAAATTTGCTCAGATGCAG 1980
QY 1981 AGCTTTGTGAGAAGTGTGCCCTCCAGTTTGAAGTGAACCCCTGACGTGACACAGTCCGC 2040
DB 1981 AGCTTTGTGAGAAGTGTGCCCTCCAGTTTGAAGTGAACCCCTGACGTGACACAGTCCGC 2040
QY 2041 CTGTGGTGTATGGAGCCAGGTGCAGACTGCTTCGGGCTGGACACAAACCCACCCCG 2100
DB 2041 CTGTGGTGTATGGAGCCAGGTGCAGACTGCTTCGGGCTGGACACAAACCCACCCCG 2100
QY 2101 GCTGGATCTCGGGCCATTAGCCAGGCGCCCTACTAGGTGGGTGGCTCAGCGCGC 2160
DB 2101 GCTGGATCTCGGGCCATTAGCCAGGCGCCCTACTAGGTGGGTGGCTCAGCGCGC 2160
QY 2161 ACCGCCCTGCTGCACATCTATGACAAAGTGTACCGTCCAGAGGGGTGCCCGGCTGGT 2220
DB 2161 ACCGCCCTGCTGCACATCTATGACAAAGTGTACCGTCCAGAGGGGTGCCCGGCTGGT 2220
QY 2221 GTCCCCAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2280
DB 2221 GTCCCCAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2280
QY 2281 GCCCAAGCTGAGGAACAATGGCATCTGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGT 2340

Db 461 AGGAAGTGAAGGAAGCATCAAGGGGATAGTTTTCRAAGGTGGCGCACCGAGCGGGCC 520
QY 722 TTGCTCTGAAATACCTTTCTCACAGAGGGTGGCTGGAGGCAGAAATGCTTCTGTGCCCC 781
Db 521 TAGCCCTGAAA---CGCCTCAGCAGAGGGTTCGCCGAGGCAGAAATGGCTCTGTGCCCC 577
QY 782 AGATCTCATATCAGTCAGTCAGTATGGAAGTCCCAAGGGGATGAGGCACTGCCATCAAGC 841
Db 578 AGATCTCTATATCTGTCACGAGTGGCAAGTCCCAAGGGGCGCTGGCTCTCCCGCTAAGC 637
QY 842 AGCTGAAGGAAGGGGTGTCACATGTGTTTCTGTGGGGTCAAGTTTCCACAGTGGAGG 901
Db 638 AGCTGAGAGAAAGGGGATCGTGTGTTTCCGTAGAGTCCGTTTTCACAGTGGGAGC 697
QY 902 AGCTGATGCACTGGCCAGCGAGCCTAGAGGGCAGCAGTGTCTGTTGGCTGAGCAGTGG 961
Db 698 AGCTGCTCACGCTGGCCAGTCAGCCGAGGAGGACCGCATGTGCTTGGCTGACGAAGTG 757
QY 962 AGGATGCCACCAAGCGCTTTCAGCACCTCAGCAGCTCGGCCATCTGCTCAGCGCCA 1021
Db 758 AGGATGCCACCAATGGCCTCTCAGCACCTCAGCAGCTCGGCCATCTGCTCAGCGCTG 817
QY 1022 CGCCAGACTCGAGGTCCAGGCTCACCCCTGTGAGCAGCAGCCTGAGATGTCGGG 1081
Db 818 ATCCAGACTCGAGGTGGAACTCACTCTGTGAGCGGAGGAGCGCTGGAGCCGTGAGG 877
QY 1082 AGTTGCGTGAATGCCCACTGTGAGAGAGTTCGGGCGGAGCCCTTGGCGTCTGGCTG 1141
Db 878 AGCTGCTGCAATGCTTGTGCTGAGAGAGTCAAGGCAAGCAGACACTGTCTGGCTC 937
QY 1142 CACACTGTCTCTACAGCTGGAAGAGAGTGTCTTAACCCACCTGCCACCTGTGTACA 1201
Db 938 TGCCCTGTCTCTACAGCTGGAAGAGAGTGTTCAGACACACACCTGCCAACTGTACA 997
QY 1202 GGACCACTTCGCCAGGCTTGTACTCCGAGCCCTGCCAGATGAGGACACATGTGTTT 1261
Db 998 GAACCACTTCGCCAGGCTTGTACTCCGAGCCCTGCCAAATGAGGACACCTGATTC 1057
QY 1262 CAGAAGCACTGGAGCGGTACCACTGCTCTGCCCCGTGGCTTGTGGAGGAGGCTAACT 1321
Db 1058 CAGAAGGTGGATAGTACCACTGTCTCTGCCACTGGCATTCGGAGGGAAGTCACT 1117
QY 1322 GTGCCCTGAAGCTGAGCTGGAATGACAGGTGACCTCTCTTCCGCTGGAAGCTCTG 1381
Db 1118 GTCCCCGAGCTGAGCTGGAATGACAGATCGATGTCTCTCTCTCTGCTGACAGTCTG 1177
QY 1382 CGGSCACCACTCTGGAGCGCTTCTCGGGGCCAAAGTCTCTGTAAGCGGTTTTCGGG 1441
Db 1178 CAGGCACCACTTGGGGGCTTCCGGAGGGCCAAAGCGCTTGTCAAGCGCTTGTGCAGG 1237
QY 1442 CCCTGCTGAGCGAGGACTCTCGGGCCGAGTGGGTGTGGCCACATACAGCAGGAGCTGC 1501
Db 1238 CCCTGCTGAGGAGGACTCCGAGCGCGCTTGGGATAGCCAGCTTATGGCAGGAATCTAA 1297
QY 1502 TGTGCGGGTG-CCTGTGGGGAGTACAGGA-TGTGCTGTGACCTGTGCTGGAGCTCGA 1559
Db 1298 TGTGCGGGTGCGCTCTCGGGGAGTACAGCAATGTGCTGGGAGCTGATCAGGAGCTTGA 1357
QY 1560 TGGCAATTCCTTCGCTGGGGCCCACTTCAGCGGCAAGTCTGCTGGCAGCGGCAG 1619
Db 1358 CAGCATTCCTTCAGCGGTGGCCCGACCTTAACCGGAGTGGCTTGTCCAGGTGGCAGA 1417
QY 1620 GCGTGGCTTCGGGAGCGCCACAGGACAGCCAGCCAGTACAGTGTGCTGTTT 1679
Db 1418 GCAGCGCTTGGAGTGGCCAGCAGGACTGTCAGGACAGCCAGCAGAGTAGTACTTCT 1477
QY 1680 GCTACTGATGACACTCCGAGGATGAGTGTGGGGCCCGCCAGCGCTGACGCAAGGCGCG 1739
Db 1478 GCTACTGAAATCAGCTCCAGGATGAGTGTCTGGGCCAGCAGCTCAGCAAGGCTCG 1537
QY 1740 AGAGCTGCTCTCTGGTGTAGCAGTGAAGCGCTGCGGGCAGAGCTGAGGAGATCAC 1799

Db 1538 GGAGCTACTCTCTCTCGGGCGTGGCAGTGAAGATCCTGTCAGGCGGAGCTGTGTGAAGATCAC 1597
QY 1800 AGCAGCCCAAAACATGTGTATGTCTACTCGGATCTCTCAGGATCTGTTCACCAAAATCCC 1859
Db 1598 CGGTAGCCCGAAGCATGTGTATGTCTCCACACAGACCTCAGGACCTGT--CAGCCAAATCC 1655
QY 1860 TGAGCTGCGAGGGGAAGCTGTGACGCGGAGCGGCGGAGCGGTCGGGACACAAAGCCCTGGA 1919
Db 1656 AGAGCTGCGAGGAGGCTATGACGCGGAGCGGAGCGGAGCGGTCGGCAGGACAGTCACTGGA 1715
QY 1920 CCTCGCTTTCATCTGTGACACCTCTGCTCAGTAGGCGCGGAGAAATTTTCTCTCAGATGCA 1979
Db 1716 CTTGGTCTTCTCTG-TGATGCTCTGTG-CYCTGTGGAGCTGAGAACTTTGCCCAAAATGCA 1773
QY 1980 GAGCTTTGTGAGAGCTGTGCTCCCTCCAGTTTGTAGGTGAACCTCAGCTGACAGAGTCCG 2039
Db 1774 GAGCTTTCATCAGCAAAATGCAACCTCCGGTTGTATGTGAATCTCTGATGACACAAAGTTGG 1833
QY 2040 CCTG 2043
Db 1834 CTTG 1837

RESULT 4
US-09-984-245-64
; Sequence 64, Application US/09984245
; Patent No. US20020165374A1
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 87 Human Secreted Proteins
; FILE REFERENCE: P2004P1
; CURRENT APPLICATION NUMBER: US/09/984,245
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/154,707
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: PCT/US98/05311
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: US 60/041,277
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/042,344
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/041,276
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/041,281
; PRIOR FILING DATE: 1997-03-21
; PRIOR APPLICATION NUMBER: US 60/048,094
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,350
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,188
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,135
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/050,937
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,187
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,099
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,352
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,186
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,069
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,095
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,131
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,096
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: US 60/048,355
; PRIOR FILING DATE: 1997-05-30


```
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 33
; LENGTH: 3449
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-907-824-33

Query Match          2.2%; Score 73.6; DB 9; Length 3449;
Best Local Similarity 48.8%; Pred. No. 2.8e-11;
Matches 242; Conservative 0; Mismatches 239; Indels 15; Gaps 1;

Qy 473 CAGTGGACATCATGTTCTGTGATGGTCTAAACAGCGTCGGGAAAGGAGCTTTGAAA 532
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2038 CAATTGACCTGGTCTTGTGATCGATGATCCAGAGTCTTGGAGAACAGAAATTTGAGG 2097
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 533 GGTCCAAAGCATTGTCATCAGCTCTGTGAGGGTCTGAGATCAGCCCCCAGAGGGTCA 592
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2098 TCGTGAAGCAGTTTGTCTACTGGAATATATAGATTCTCTTGACAAATTTCCCCAAAGCCGCTC 2157
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 593 GAGTGGGAGCATTCAGTTTCAGTTCCACTCTCTCATCTGGAATTCCTCTTGATTCATTTT 652
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2158 GAGTGGGCTGCTCCAGTATTCACACAGGTCACACAGAGTTCACTCTGAAACTTCA 2217
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 653 CAACCCAAACAGAGTGAAGGCAAGAATCAAGAGATGGTTTTCAAGAGGGGGCGCACGG 712
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2218 ACTCAGCCAAAGACATGAAAGGCGGTGGCCACATGAAATACATGGAAGAGGCTCTA 2277
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 713 AGAGGAACTTCTCTGAAATACCTTCTGCACAGAGGTTCCCTGAGGCGCAAAATG--- 769
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2278 TGACTGGGCTGGCCCTGAAACACATGTTTGAGAGAAGTTTACCCAAAGGAGGGGCCA 2337
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 770 -----CTTCTGTGCCAGATCCCTCATCATCTGTCAGTGGGAAGTCCCAAGG 817
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2338 GGCCCTTTTCCACAGGGTGGCCAGAGAGCCATTTGTTTCCAGCAGCGGCTCAGG 2397
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 818 GGGATGGGCACTGCCATCCAAAGCAGCTGAAGGAAAGGGGTGCTCACTGTGTTGCTGG 877
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2398 ATGAGCTCTCGAGTGGGCCAGTAAGCCAAAGGCCAATGGTATCACTATGTATGCTGTG 2457
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 878 GGGTCAGGTTTCCAGGTGGGAGGCTGATGATGCTGCGCAGCGAGCTAGAGGGCAGC 937
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2518 ATCTCTTCTATGCCGA 2533
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-09-907-841-33
; Sequence 33, Application US/09907841
; Publication No. US20020198366A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
```

Db 2458 GGGTAGGAAACCCATTGAGGAGGAACTACAAGAGATTGCTTCTGAGCCCAACAAAGC 2517
QY 938 ACGTGCTGTTGGCTGA 953
Db 2518 ATCTCTTCTATGCCGA 2533

RESULT 13

US-09-904-011-33
; Sequence 33, Application US/09904011
; Publication No. US20030003530A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,011
; PRIOR FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 33
; LENGTH: 3449
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-904-011-33

Query Match 2.2%; Score 73.6; DB 9; Length 3449;
Best Local Similarity 48.8%; Pred. No. 2.8e-11;
Matches 242; Conservative 0; Mismatches 239; Indels 15; Gaps 1;
QY 473 CAGTGGACATCATGTTCTGTTAGATGGTCTTAACAGCGCTCGGAAAGGAGCGCTTTGAAA 532
Db 2038 CAATTGACCTGCTCTTTGTGATCGATGATCCAAAGAGTCTTGGAAGAAGAGAATTTTGGG 2097
QY 533 GGTCCAAGCACATTGGCATCACAGTCTGTGACGGTCTGCACATCAGCCCGGAGAGGTCA 592
Db 2098 TCGTGAAGCAGTTTGTCACTGGAATTTATAGATTCCTTGACAATTTCCCGCCAAAGCGCTC 2157
QY 593 GAGTGGGAGCATTCAGTTTCAGTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 652
Db 2158 GAGTGGGCTGCTCCAGTATTCACACACAGGTCCACACAGATTCACCTCTGAGAACTTCA 2217
QY 653 CAACCCACAGAGTGAAGGCAAGATCAAGAGGATGCTTTTCAAGAGGAGGCGGCACGG 712
Db 2218 ACTCAGCCAAAGACATGAAAAAGCGGTGGCCCATACATGATGGAAGAGGCGCTCA 2277
QY 713 AGACGGAACTTGTCTGAAATACCTTCTCAGACAGGGTGTGCTGAGGCGAGAAATG--- 769
Db 2278 TCAGTGGGCTGCGCCCTGAAACACATGTTTGAGAGAAGTTTACCACAGGAGAGGGGCCA 2337
QY 770 -----CTTCTGTGCCCCAGATCCTCATCATCCTCCTCCTCCTCCTCCTCCTCCTCCT 817
Db 2338 GGCCCTCTTTCCCAAGAGGTGCCAGAGAGCCATTGTGTTCCACGCGACGCGCTCAGG 2397
QY 818 GGGATGTGGCACTGCCATCCAAAGCAGCTCAAGGAAAGGGGTGTCATGTGTTGTGTGG 877
Db 2398 ATGAGCTCTCCAGTGGGCGCCAGTAAGCCAGGCCAATGGTATCATCTATGATCTCTGTG 2457
QY 878 GGGTCAGGTTTCCAGGTTGGGAGGAGCTCATGCACTGGCCAGCGAGCGCTAGAGGCGACG 937
Db 2458 GGGTAGGAAAGCCATTGAGGAGGAACTACAAGAGATTGCCTCTGAGCCCAACAAAGC 2517
QY 938 ACGTCTGTTGGCTGA 953
Db 2518 ATCTCTTCTATGCCGA 2533

RESULT 14

US-10-028-072-293
; Sequence 293, Application US/10028072
; Publication No. US20030004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel

APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang
TITLE OF INVENTION:
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/028,072
CURRENT FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: 60/049911
PRIOR FILING DATE: 1997-06-18
PRIOR APPLICATION NUMBER: 60/056974
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059115
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059117
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059122
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059184
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059352
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059588
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059836
PRIOR FILING DATE: 1997-09-24
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062285
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062287
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062814
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/062816
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063045
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063082
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/063127
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063327
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063329
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063550
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063561
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063704
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063733
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063735
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063738
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063755
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064248
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/064809
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065846
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21

PRIOR APPLICATION NUMBER: 60/066453
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066511
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069212
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069278
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069334
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069694
PRIOR FILING DATE: 1997-12-16
PRIOR APPLICATION NUMBER: 60/072320
PRIOR FILING DATE: 1998-01-23
PRIOR APPLICATION NUMBER: 60/073612
PRIOR FILING DATE: 1998-02-04
PRIOR APPLICATION NUMBER: 60/074086
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/074092
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-02-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081695
PRIOR FILING DATE: 1998-04-14
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081818
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082999
PRIOR FILING DATE: 1998-04-24
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085149
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085697
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086414
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086430


```

; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
;
; SEQ ID NO 33
; LENGTH: 3449
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-906-742-33

```

Query Match	2.28;	Score	73.6;	DB	9;	Length	3449;
Best Local Similarity	48.8%;	Pred. No.	2.8e-11;				
Matches	242;	Conservative	0;	Mismatches	239;	Indels	15;
Gaps	1;						

QY	473	CAGTGGACATCATGTTTCTCTTAGATGGGTCCTACACCGTCGGGAAGGGAGCTTCGAAA	532
Db	2038	CAATTGACCTGGCTTTGTGATCGATGGATCCGAAGACTCTGGAGAAGAAATTTTTCAGG	2097
QY	533	GGTCCAAAGCATTTGGCCATCACAGTCTGTGACGCTCTGGACATCAGCCCCGAGAGGCTCA	592
Db	2098	TCGTGAAGCAGTTTGTCCACTGGAAATATAGATTCTTGACAATTTCCCCCAAGCCGCTC	2157
QY	593	GAGTGGAGCATTTCCAGTTCCAGTTCCACTCCTCATCTGGGAATCCCCTTGGATTCATTTT	652
Db	2158	GAGTGGGGCTGCCTCCAGTATTTCCACACAGGTCACACAGAGTTTCACCTCTGAGAAACTCA	2217
QY	653	CAACCCACAGGAGAGTCAAGGCCAAGATCAAGAGGATGTTTTTCAAAGGAGGGCGCACGG	712
Db	2218	ACTAGCCAAAGACATGAAAAAGCCGTGCCCCACATGAAATACATGGGAAGGGCTCTA	2277
QY	713	AGACGGAACTTGCTCTGAAATACCTTCTGCACAGAGGTTTGCCTGGAGCGAGAATG---	769
Db	2278	TGACTGGCTGGCCCTGAAACACATGTTTCGAGAGAAGTTTATCCCAAGAGAGAGGGGCCA	2337
QY	770	-----CTTCTGTGCCCCAGATCCTCATCATCTGTCACGTATGCGGAAGTCCCAGG	817
Db	2338	GGCCCCCTTTCCCAAGGGTGCCCAAGCAGCATTTGTGTTCCCGACGAGCGGCTCAGG	2397
QY	818	GGGATGTGGCACTGCCATCCACAGCAGCTGAAGGNAAGGGGTGTCACCTGTGTTTGTCTGG	877
Db	2398	ATGACGTCTCCGAGTGGCCAGTAAAGCCAAATGGTATCATCTATGTATGCTCTGTG	2457
QY	878	GGGTACAGTTTCCAGGTGGGAGGAGCTGCTACCTGCGACCTGGCCAGCGAGCCTAGAGGGCAGC	937
Db	2458	GGGTAGGAAAAGCCATTGAGGAGGAACTACAGAGATTGGCTCTGAGCCCAACAACAGC	2517
QY	938	ACGTGCTGTTGGGTGA	953
Db	2518	ATCTCTCTATGCGGA	2533

Search completed: May 5, 2003, 04:59:36
Job time : 346 secs

7

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 01:38:18 ; Search time 3008 Seconds
(without alignments)
18171.465 Million cell updates/sec

Title: US-09-930-020A-1
Perfect score: 3375
Sequence: 1 gacagtgttcggtgcac.....tgtccaccttgaagtcttc 3375

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_nam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	458.8	13.6	1024	12	BF164945	BF164945 PM4-EN0006
C	2	428.6	12.7	535	12	BF849816 PM4-EN0006
3	393	11.6	664	10	BB617396	BB617396 BB617396
C	4	389.8	11.5	632	17	AZ344444 AZ344444
5	349.4	10.4	638	10	BB613951	BB613951 BB613951
C	6	340	10.1	406	10	AW206492 UI-H-B11-

7	303.8	9.0.	647	10	BB655280	
8	294.8	8.7	631	10	BB612001	
9	264.2	7.8	420	10	BB846685	
10	219	6.5	393	12	BF606846	
C	11	208.6	6.2	282	10	BE143292
12	176.4	5.2	629	13	BJ003847	
C	13	125.2	3.7	201	10	AW856235
14	114.8	3.4	438	13	BJ036168	
C	15	113	3.3	723	14	BM702190
16	98.8	2.9	591	10	AV593500	
C	17	92.4	2.7	483	17	AZ859528
18	88.2	2.6	1011	14	BQ934713	
19	88.2	2.6	1993	11	AK014342	
20	85.4	2.5	615	13	BJ499393	
21	83.4	2.5	567	10	AW855319	
22	83.2	2.5	402	13	BJ067406	
23	75.4	2.2	563	9	AI385774	
24	75.2	2.2	514	13	BI067543	
25	74.4	2.2	623	13	BJ501694	
26	73.4	2.2	510	13	BI337438	
27	72.8	2.2	742	13	BM050442	
28	72.2	2.1	600	10	BB577711	
29	72.2	2.1	716	13	BI658919	
30	72.2	2.1	1149	13	BM454041	
31	72.2	2.1	1407	11	AK017553	
32	72	2.1	867	14	BQ876319	
33	71.4	2.1	580	10	AV594887	
34	71.4	2.1	603	10	AV592803	
35	71.2	2.1	654	10	BB617384	
36	71	2.1	558	10	AV599656	
37	70.6	2.1	568	10	AV607535	
38	70.6	2.1	590	10	BB616450	
39	70.6	2.1	798	13	BI546662	
40	70.2	2.1	567	10	AV595999	
41	69.8	2.1	496	17	AQ202422	
42	69.6	2.1	966	9	AL548954	
C	43	69	536	10	BE048895	
44	69	2.0	585	13	BM313579	
45	69	2.0	618	14	BQ778029	

ALIGNMENTS

RESULT 1
BF164945
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

BF164945 1024 bp mRNA linear EST 30-OCT-2000
601778179F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:4019899 5',
mRNA sequence.
BF164945
BF164945.1 GI:11045297
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1024)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue procurement: Gilbert Smith, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LAM9272 row: m column: 20
High quality sequence stop: 667.
Location/Qualifiers
1..1024


```

Db 588 AGCTGAGAGAAAGGGGCACTGCTGGTGTGGCCGTAGGAGTCCGTTTCCAGGTGGGAGG 647
QY 902 AGCTGCATCACTGGCC 918
      |||||  |||||
Db 648 AGCTGCTCAGCTGGCC 664

RESULT 4
LOCUS      A2344444/c
DEFINITION 1M00782P24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M00782P24 F, DNA sequence.
ACCESSION  A2344444
VERSION    A2344444
KEYWORDS   GSS.
SOURCE     A2344444.1 GI:10423680
           house mouse.
ORGANISM   Mus musculus.
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
           1 (bases 1 to 632)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0078 row: P column: 24
Seq primer: CTTGTAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 632.
Location/Qualifiers
1. 632
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0078P24"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 134 a 199 c 178 g 120 t
ORIGIN

Query Match 11.5%; Score 389.8; DB 17; Length 632;
Best Local Similarity 83.1%; Pred. No. 1.1e-83;

```

```

Matches 456; Conservative 0; Mismatches 92; Indels 1; Gaps 1;
QY 1894 CCAGGGTCCGGACACAAAGCCCTGGACCTCGCTCTTCATGTTGGACACCTCTCTGCTCAGTA 1953
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 565 CCAGGCTCCAGGCACAGTCACCTGGACCTGGTCTTCTCTGTTGATGCTCTGCTCTG 506
QY 1954 GGGCCCGAGAAATTTTGCTCAGATGAGAGCTTTGTGAGAAAGCTGTGCCCTCCAGTTTGAG 2013
      ||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 505 GGACGTGAGAACTTTGCCCAATGCAGAGCTTCATCAGGAAATGCACCTCCGGTTGAT 446
QY 2014 GTGACCCCTGAGTGCACAGCTGGCCCTGGTGTATGTCAGCAGCAGGTGAGAGTGGC 2073
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 445 GTGAATCCTGATGTGACAGGTGGCCCTGGTGGTATATGGCAGCGGGGTGCAGACTGCC 386
QY 2074 TTCGGGCTGGACACCAAAAC-CCACCCGGGCTGCGATGCTGCGGGGCATTAGCCAGGCCCC 2132
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 385 TTCGGGCTAGACACCAATCCACCCGGGCTGCCGTGTTGGCGGCATGAGCCAGGCTCC 326
QY 2133 CTACCTAGTGGGGTGGGTCTAGCCGGCACCCGCTCTGTCACATCTATGACAAAGTGAT 2192
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 325 CTACCTAGGGGTGTGGGTTCCTGCTGGTACAGCTTGTCTGCACATTTGAGGACAAAGTGAT 266
QY 2193 GACCGTCCAGAGGGGTGGCCGCTGCTGCCAAAGCTGTGGTGTCTCACAGGCGG 2252
      ||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 265 GACGTCCAGAGGGGTGGCCGCTGCTGCCAAAGCTGTGGTGTCTCACAGGCGG 206
QY 2253 GAGAGCGCAGAGGTGACAGCGGTTCTCTCCAGAGAGCTGAGGAAACATCTCTCT 2312
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 205 GAGTGGGCAGAAATGTCAGCTGTCCCTGCCAGAGCTGAGAGCAATGGCATCTCAT 146
QY 2313 CTTGGTGTGGGCGTGGGCCCTGCTCTAGTGAAGGTCTGCGGAGGCTTGCAGGTCGCCG 2372
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 145 CTTGGTATGAGCTGGGAGCGCTCTCTCAGGAGGACGATACGAGGCTTTCGCGGTCCACG 86
QY 2373 GGATTCCCTGATCCAGTGGCGAGCTTACCCTGCGGTACCCACGAGAGCTGTCTCAT 2432
      ||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 85 CGACTCCCTGATCCATGTCGGAGCTTACACAGACCTTCCTGACCATCAGACATGCTCAT 26
QY 2433 TGAGTGGCT 2441
      |||||
Db 25 CGAGTGGCT 17

RESULT 5
BB613951
LOCUS      BB613951
DEFINITION BB613951 RIKEN full-length enriched, 0 day neonate head Mus
musculus cDNA clone 4832416E03 5', mRNA sequence.
ACCESSION  BB613951
VERSION    BB613951
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus.
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
           1 (bases 1 to 638)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiranoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp.

```



```

NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2,
NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_CLL1, NCI_CGAP_Lei2,
NCI_CGAP_Brn23, NCI_CGAP_Lu5, NCI_CGAP_Lu24,
NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,
NCI_CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI_CGAP_Kid3 pool 1 LLM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonoids 1322376-1323911, 1456008-1456775
, 1500552-1502855); NCI_CGAP_Kid5 pool 1 LLM 3338-3342
, 3722-3725, 3776-3778 (IMAGE Clonoids 1323912-1325831,
1471368-1472903, 1492104-1493255); NCI_CGAP_Lu5 pool 1
LLM 3575-3582, 3851-3854 (IMAGE Clonoids 1414920-1417991,
1520904-1522439); NCI_CGAP_GC4 pool 1 LLM 3164-3167,
3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631,
1469064-1470983, 1475592-1476743); NCI_CGAP_Pr22 pool 1
LLM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids
985608-986759, 1101192-1101959, 1217928-1220615);
NCI_CGAP_Co10 pool 1 LLM 2844-2653, 2871-2872 (IMAGE
Clonoids 1057416-1061255, 1144584-1145351). Subtraction
was performed as previously described [Bonaldi, Lennon &
Soares (1996): Normalization and Subtraction: Two
Approaches To Facilitate Gene Discovery. Genome Research
6, 791-806.
TAG_LIB=NCI_CGAP_Co10
TAG_TISSUE=colon
TAG_SEQ=AAACG"
BASE COUNT 80 a 112 c 107 g 107 t
ORIGIN

Query Match 10.1%; Score 340; DB 10; Length 406;
Best Local Similarity 90.7%; Pred. No. 1.le-71;
Matches 390; Conservative 0; Mismatches 0; Indels 40; Gaps 1;

QY 2482 AGCCCGTGCATGAATGAGGCGAGCTCGTCCTGCAGATGGAGCTACCGCTGCAAGTGT 2541
DB 406 AGCCCGTGCATGAATGAGGCGAGCTCGTCCTGCAGATGGAGCTACCGCTGCAAGTGT 347
QY 2542 CGGGATGGCTGGAGGGCCCCCACTGCGAGAACCGTGTGAGTGAGCTCTTGTCTGTATGT 2601
DB 346 CGGGATGGCTGGAGGGCCCCCACTGCGAGAAC----- 313
QY 2602 GTGAGCCGATGATGATTTCTTGAGAGCCCTTGAGGCACATGGCTCCCTGTCAGGAGGCG 2661
DB 312 -----GATTTCTGAGAGCCCTTGAGGCACATGGCTCCCTGTCAGGAGGCG 267
QY 2662 AGCAGCCGTACCCCTCCAGCAACTACAGAGAGGCGCTGGGCACCTGAAATGGTGCCTACC 2721
DB 266 AGCAGCCGTACCCCTCCAGCAACTACAGAGAGGCGCTGGGCACCTGAAATGGTGCCTACC 207
QY 2722 TTCTGGAATGTCTGTGCCCGAGTCTTGAATGTCTGTCCCGCGTGGCCAGGACCA 2781
DB 206 TTCTGGAATGTCTGTGCCCGAGTCTTGAATGTCTGTCCCGCGTGGCCAGGACCA 147
QY 2782 CTATTTCTACTGAGGAGGAGGTGTCCTCACTGAGCCATGCTGCTTAGACACAGAAA 2841
DB 146 CTATTTCTACTGAGGAGGAGGTGTCCTCACTGAGCCATGCTGCTTAGACACAGAAA 87
QY 2842 CGAGCTGATGTACCCACAAACGATGTTGTTGAAAAGTTTGTATGTGAAGTAATACCC 2901
DB 86 CGAGCTGATGTACCCACAAACGATGTTGTTGAAAAGTTTGTATGTGAAGTAATACCC 27
QY 2902 ACTTCTGTGA 2911
DB 26 ACTTCTGTGA 17

RESULT 7
BB655280
LOCUS
DEFINITION
BB655280 RIKEN full-length enriched, 9 days embryo Mus musculus
CDNA clone D030048B17 5', mRNA sequence.

```

ACCESSION VERSION KEYWORDS SOURCE ORGANISM

REFERENCE AUTHORS

TITLE JOURNAL COMMENT

BB655280 1 GI:16489108
BB655280.1
EST.
Mus musculus
house mouse
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 647)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
Hiramoto,K., Hori,F., Ishi,Y., Ito,M., Kawai,J., Konno,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.

FEATURES source

Location/Qualifiers
1..647
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D030048B17"
/clone_lib="RIKEN full-length enriched, 9 days embryo"
/dev_stage="9 days embryo"
/lab_host="DH10B"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCCGACCTCGAGTTTCTTTTCTTTT 3'], cDNA was
prepared by using trehalase thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGATTCGAGTTAATTAATATCCCTCCCTCCCTCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified

BOOKNAME
COMMENT
published (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic

FLC I ⁿ	modified pBluescript KS(+)	after bulk excision from Lambda
124 a	193 c	129 t
		3 others

SCGGGGCGTCCGACCCAGGCTAACACATGCCCTCCACTTCTGCTTCTACCGGCATCTACA 207

TTTTCCTGTTTTCCAGAGTCCCCCATCTCTCCCTCTCCAGGAAGTCCATGTAAAGCAAAG 421

Query Match	302	302
Best Local	148	148
Matches	362	362

KEYWORDS	EST.	ACCESSION	
SOURCE	Bos taurus	BE143292.1	
ORGANISM	cow.	GI:8606013	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.	EST.	
AUTHORS	Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Perle, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keefe, J.W.	human	
TITLE	Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle	Homo sapiens	
JOURNAL	Genome Res. 11 (4), 626-630 (2001)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
MEDLINE	21180013	1 (bases 1 to 282)	
COMMENT	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smith@mail.marc.usda.gov Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options. PCR Primers FORWARD: AGGAACAGCTATGACCAT BACKWARD: GTTTCCTCCAGTCACGACG Plate: 57 row: K column: 1 Seq primer: ATTAGGTGACACTATAG. Location/Qualifiers 1. .393 /organism="Bos taurus" /db_xref="taxon:9913" /clone_lib="MARC 3BOV" /tissue_type="pooled" /lab_host="DH10B" /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI; Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle." 64 a 130 c 123 g 76 t	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) 20202663 Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=st2-MR0-HT0161-221099-002-c08st3-1999-10-22st4-1) Seq primer: puc 18 forward High quality sequence start: 8 High quality sequence stop: 282. Location/Qualifiers 1. .282 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="HT0161" /dev_stage="Adult" /note="Organ: head-neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." 71 a 78 c 63 g 70 t	
FEATURES	source	FEATURES	
BASE COUNT	64 a 130 c 123 g 76 t	BASE COUNT 71 a 78 c 63 g 70 t	
ORIGIN		ORIGIN	
Query Match	6.5%; Score 219; DB 12; Length 393;	Query Match 6.2%; Score 208.6; DB 10; Length 282;	
Best Local Similarity	84.5%; Pred. No. 2.8e-42;	Best Local Similarity 93.7%; Pred. No. 7.8e-40;	
Matches	246; Conservative 0; Mismatches 45; Indels 0; Gaps 0;	Matches 239; Conservative 0; Mismatches 14; Indels 2; Gaps 2;	
QY 1895	CAGGGTCCGACACAGCCCTGGACCTCGTCTTCATGTTGGACACCTCTGCCTCAGTAG 1954	QY 3026	CGCCAGAATGTTGTGACACAGTAATGCCACAGAGGCCCTTTACTAGACATCCTTTGG 3085
Db 103	CAGGTGCCAGGCCAGTCGCTGGACCTCGTCTTCATGCTGGAGCGCTCGGTCTCAGTGG 162	Db 282	CGCCAGAATGTTGTGACACAGTAATGCCACAGAGGCCCTTTACTAGACATCCTTTGG 223
QY 1955	GGCCCGAGAAATTTGCTCAGATCAGAGCTTTGTGAGAAGCTGTGCCCTCCAGTTTGGG 2014	QY 3086	ACGGCGAAGGCCACGCCCTTTCAAGATGGAAGCAGCAGCTTTTCCACTTCCCAGAGAC 3145
Db 163	GGCCCGAGAAATTTGCTCAGATCAGAGCTTTGTGAGAAGCTGTGCCCTCCAGTTTGGG 222	Db 222	ACGGCGAAGGCCACGCCCTTTCAAGATGGAAGCAGCAGCTTTTCCACTTCCCAGAGAC 163
QY 2015	TGAACCTTGACGTGACACAGGTGGCCCTGGTGGTGTATGGCAGCCAGGTGCAGACTGCCT 2074	QY 3146	ATTCTCG-ATGCATTGTCATTGAGTCTCTGAAGGGGGCTTGAGGAGCGCTTTGTGACTTCTT 3204
Db 223	TGAACCCGATGTGACGAGATCGGTCTGGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 282	Db 162	ATTCTGGAATGCAATTTGCAATTTGAGTCTGAAGGGGGCTTGAGGAGCGCTTTGTGACTTC-T 104
QY 2075	TCGGGTGGACACCAACCCAGCCGCTCGATGCTGCGGGCCATTAGCCAGGCCCTT 2134	QY 3205	GGCGACTGCCTTTTGTGTGTGGAAGAGACTTGGAAAGGCTCTCAGACTGAATGTGACCAAT 3264
Db 283	TCGGGTGGACACCCACCTACGCGTGTGTGAGTGTGTGCGGGCTCTGAGCCAGGCCCTT 342	Db 103	GGCGACTGCCTTTTGTGTGTGGAAGAGACTTGGAAAGGCTCTCAGACTGAATGTGACCA 44
QY 2135	ACCTAGGTGGGTGGCTCAGCGCCAGCCGCTTCTGCACATCTATGACA 2185	QY 3265	TAAACAGCTTGGTTG 3279
Db 343	ACCTGGAGGGGTGGCTCAGCAGCGACCGGACTTGTCTGCACATCTATGACA 393	Db 43	TACCAGCATTGATTG 29
RESULT 11			
BE143292/c			
LOCUS			
DEFINITION	MR0-HT0161-221099-002-c08 HT0161 Homo sapiens cDNA, mRNA sequence.		

Df 527 AAATACATCTTGAGGAAGGGTTTTCCGGGTGCGGTAACACTTCTCCAGCAGTCGCCGCAT 586

Qy 786 CCTCATCATGCTCACTGATGGGAAGTCCCAGGGGATGTGGC 827
| |||| | | | | | | | | | | | | | | | |
Df 587 GCCCATCTCTGTCAGATGGGAGGTCTTCAGGGCAGCGTGGC 628

RESULT 13
AW856235/c
LOCUS AW856235 linear EST 19-MAY-2000
DEFINITION RC1-Ct0286-050400-018-e03 Ct0286 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW856235
VERSION AW856235.1 GI:7951928
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 201)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,C.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL 20202663
MEDLINE
COMMENT Laboratory of Cancer Genetics
Contact: Simpson A.J.G.
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LIICR Human Cancer Genome
Project. This entry can be seen in the following URL
<http://www.ludwig.org.br/scripts/gethtml2.pl?ti=et5-RC1-Ct0286-050400-018-e03&t3=2000-04-05&t4=1>

Seq primer: puc 18 forward
High quality sequence start: 26
High quality sequence stop: 200.
Location/Qualifiers
1..201
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0286"
/dev_stage="Adult"

/note="Organ: colon; Vector: puci18; Site_1: Sma1; Site_2:
Sma1; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the PUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 40 a 69 c 54 g 38 t
ORIGIN

Query Match 3.7%; Score 125.2; DB 10; Length 201;
Best Local Similarity 87.9%; Pred. No. 1.2e-19;
Matches 160; Conservative 0; Mismatches 18; Indels 4; Gaps 2;

Qy 2241 GCTCACAGCGGGAGAGCGGCAGAGGATGCAGCCGTTCTCTCCCCAGAAAGCTTGAGGAACA 2300
| |||| | | | | | | | | | | | | | | | |
Df 193 GTTCGAGCGGGAGGGCGCAGAGGATGCAGCCGTTCTCTCACCAAAGAGCTTGAGGAACA 134
| |||| | | | | | | | | | | | | | | | |

Qy 2301 TGGCATCTCTTGTCGTGGCGTGGCGCTCTCTAAGTAGGCTCTCGGAGGCT 2360
| |||| | | | | | | | | | | | | | | | |
Df 133 TGCCATCTCTCTTTGGTCGT-GGCGTGGGGCCTGCTCTAAGTAGGCTCTCGGAGGCT 75
| |||| | | | | | | | | | | | | | | | |

Ov 2361 TGCAGGTCGCCGGATTCCCTGATCCAGCTGGCAGCTTTACGCCACCTTCGCGTACCACCA 2420
| |||| | | | | | | | | | | | | | | | |

```

RESULT 15
BM702190/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

BM702190 723 bp mRNA linear EST 28-FEB-2002
UI-E-CQ1-aeY-m-08-0-UI.r1 UI-E-CQ1 Homo sapiens cDNA clone
UI-E-CQ1-aeY-m-08-0-UI 5', mRNA sequence.
BM702190
BM702190.1 GI:19015448
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 723)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery.
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 353-608, >LINE2
Seq primer: M13 Reverse.
Location/Qualifiers
1..723
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-CQ1-aeY-m-08-0-UI"
/clone_lib="UI-E-CQ1"
/tissue_type="optic nerve"
/dev_stage="adult"
/lab_host="DHI08 (Life Technologies) (T1 phage resistant)"
/notes="Organ: eye; Vector: pT73-pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-CQ1 is a normalized cDNA library containing the
following tissue(s): optic nerve. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
CCATTAAAGT. This library was created for the program, Gene
Discovery in the Visual System, supported by National Eye
Institute (NEI)."
```

BASE COUNT	147 a	241 c	142 g	190 t	3 others
ORIGIN					
Query Match	3.3%	Score 113;	DB 14;	Length 723;	
Best Local Similarity	100.0%	Pred. No. 2.2e-16;			
Matches 113;	Conservative	0;	Mismatches	0;	Gaps 0;
Indels	0;				
QY 2615	GGATCTTGAGAGCCCTCCTGAGGCACATGCTCCCTGTCAGGAGGCGCAGACGGTACCC	2674			
Db 113	GGATCTTGAGAGCCCTCCTGAGGCACATGCTCCCTGTCAGGAGGCGCAGACGGTACCC	54			
OY 2675	CTCCAGCAACTACAGAGAAGGCGCTGGGCACTGGAATGGTGCCTACCTCTCGG	2727			

Db 53 CCCCCAGCACTACAGAGAGGCGCTGGCACTGAAATGGTGCGCTACCTTCGG 1

Search completed: May 5, 2003, 04:53:50
Job time : 3030 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 04:52:36 ; Search time 5744 Seconds
(without alignments)
17099.925 Million cell updates/sec

Title: US-09-930-020A-1

Perfect score: 3375

Sequence: 1 gacagtgttcgcggtgcac.....tgtccaccttgaagtgcttc 3375

Scoring table: OLIGO.NUC

Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_hgt:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pi:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_ind:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pi:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_hgt_hum:*

31: em_hgt_inv:*

32: em_hgt_other:*

33: em_hgt_mus:*

34: em_hgt_pln:*

35: em_hgt_rtd:*

36: em_hgt_nam:*

37: em_hgt_vrt:*

38: em_sy:*

39: em_hgtgo_hum:*

40: em_hgtgo_mus:*

41: em_hgtgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	589	17.5	123110	9	AC005383	Homo sapi
C 2	316	9.4	160420	9	AC022023	Homo sapi
C 3	264	7.8	136357	2	AL161942	Homo sapi
4	69	2.0	496	11	G50805	SHGC-B3858
5	31	0.9	154748	2	AC12150	Mus muscu
6	25	0.7	1728	9	F243081S03	Homo sapi
7	25	0.7	11344	9	AF034611	Homo sapi
8	25	0.7	147280	9	AC067747	Homo sapi
9	25	0.7	152895	2	AC119647	Homo sapi
C 10	23	0.7	622	9	HSNOTCH2	H. sapiens N
11	23	0.7	789	9	HSU77493	Human Notch
12	23	0.7	924	9	HS296168	Homo sapi
13	23	0.7	1099	9	HS406936	Homo sapi
14	23	0.7	9722	9	AF308601	Homo sapi
15	23	0.7	9723	6	AR102329	Sequence
16	23	0.7	9723	6	AX357059	Sequence
17	23	0.7	9723	6	AX417417	Sequence
18	23	0.7	11189	9	AF315356	Homo sapi
19	23	0.7	109612	9	AL512503	Human DNA
C 20	23	0.7	142565	10	AL592545	Mouse DNA
21	23	0.7	161987	9	AC006070	Homo sapi
22	23	0.7	174032	2	AC025904	Homo sapi
C 23	23	0.7	180836	2	AC025241	Homo sapi
C 24	23	0.7	182571	2	AC095873	Rattus no
25	23	0.7	197658	2	AL646094	Mus muscu
C 26	23	0.7	347572	2	AC107303	Homo sapi
27	22	0.7	56030	2	AC129682	Rattus no
C 28	22	0.7	90335	2	AP003917	Oryza sat
29	22	0.7	152895	2	AC119647	Rattus no
C 30	22	0.7	161987	9	AC006070	Homo sapi
C 31	22	0.7	174032	2	AC025904	Homo sapi
C 32	22	0.7	188718	2	AC113703	Rattus no
C 33	22	0.7	201802	2	AC124473	Mus muscu
C 34	22	0.7	264009	2	AC115294	Mus muscu
C 35	22	0.7	318085	2	AC127297	Mus muscu
36	21	0.6	1100	5	AF300613	Gallus ga
37	21	0.6	1182	4	AF304108	Mesoplo
38	21	0.6	1730	9	AB063081	Macaca fa
C 39	21	0.6	3024	9	AK055872	Homo sapi
C 40	21	0.6	4865	9	HSHP201	Human hp2 9
C 41	21	0.6	5421	9	CHPAPCC1	Chimpanzee
42	21	0.6	7319	10	AF164486	Rattus no
C 43	21	0.6	11551	9	HUMHPARS1	Human hapto
C 44	21	0.6	38542	9	HUMHP2HPR	Human hapto
45	21	0.6	46638	9	AC104823	Homo sapi

ALIGNMENTS

RESULT 1

AC005383

LOCUS

DEFINITION

AC005383

AC005383

AC005383.1

HTG.

SOURCE

ORGANISM

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 123110)

Smith, D.R.

Sequencing of Human Chromosome 10

AC005383 123110 bp DNA linear FRI 31-OCT-1998
Homo sapiens chromosome 10 clone CIT987SK-1144G6 maf 10q25.1,
complete sequence.

AC005383.1 GI:3818355

HTG.

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 123110)

Smith, D.R.

Sequencing of Human Chromosome 10

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 123110)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (05-AUG-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA
REFERENCE 3 (bases 1 to 123110)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA
COMMENT On Oct 31, 1998 this sequence version replaced gi:3808081.
FEATURES
source
BASE COUNT 31394 a 28668 c 29634 g 33413 t 1 others
ORIGIN
Query Match 17.5%; Score 589; DB 9; Length 123110;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2615 GGATTCCTGAGCGCCCTGAGGCACATGGCTCCCGTGCAGGAGGCGCAGCGCTACCC 2674
DB 78153 GGATTCCTGAGCGCCCTGAGGCACATGGCTCCCGTGCAGGAGGCGCAGCGCTACCC 78212
QY 2675 CTCACGCACTACAGAGAGCGCTGGGCACTGAATGGTCTTCTGGAATGCT 2734
DB 78213 CTCACGCACTACAGAGAGCGCTGGGCACTGAATGGTCTTCTGGAATGCT 78272
QY 2735 GTGCCCGAGGCTCTTGAATGCTGTCTCCCGCGTGGCGAGCACCACTATTCTCACTCA 2794
DB 78273 GTGCCCGAGGCTCTTGAATGCTGTCTCCCGCGTGGCGAGCACCACTATTCTCACTCA 78332
QY 2795 GGGAGGAGGATGCCAACTGAGGCATGCTGCTTAGAGACAAAGAGCGCTGATGCTCA 2854
DB 78333 GGGAGGAGGATGCCAACTGAGGCATGCTGCTTAGAGACAAAGAGCGCTGATGCTCA 78392
QY 2855 CCCACAAAGCGTGTCTGAAAGTTTTCATGTTGAAGTAATACCCACTTTCTGTACCT 2914
DB 78393 CCCACAAAGCGTGTCTGAAAGTTTTCATGTTGAAGTAATACCCACTTTCTGTACCT 78452
QY 2915 GCTGTGCTTGTGAGGCTATGTCATCTGCCACCTTTCCCTTGAGGATAAACAAGGGGTC 2974
DB 78453 GCTGTGCTTGTGAGGCTATGTCATCTGCCACCTTTCCCTTGAGGATAAACAAGGGGTC 78512
QY 2975 CTGAAGACTTAATTTAGCGGCTGACGTTCTTTGCAACAAATCAATGCTGCCAGAAT 3034
DB 78513 CTGAAGACTTAATTTAGCGGCTGACGTTCTTTGCAACAAATCAATGCTGCCAGAAT 78572
QY 3035 GTTGTGTACACAGTAATGCCAGCAGCGCTTACTAGAGCATCTTTGGAGGGGGAAG 3094
DB 78573 GTTGTGTACACAGTAATGCCAGCAGCGCTTACTAGAGCATCTTTGGAGGGGGAAG 78632
QY 3095 GCCACGGCTTTCAAGATGGAAGAGCAGCAGCTTTCCACTTCCCGAGACATTTCTGGAT 3154
DB 78633 GCCACGGCTTTCAAGATGGAAGAGCAGCAGCTTTCCACTTCCCGAGACATTTCTGGAT 78692
QY 3155 GCATTTGCATGAGTCTGAAAGGGGCTTTGAGGAGCGTTTGTGACTTCT 3203
DB 78693 GCATTTGCATGAGTCTGAAAGGGGCTTTGAGGAGCGTTTGTGACTTCT 78741

RESULT 2
AC022023/c
LOCUS AC022023
DEFINITION Homo sapiens chromosome 10 clone RP11-411P18, complete sequence.
AC022023
VERSION AC022023.7 GI:20901843

KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 160420)
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome Sequence Data
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 160420)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2000) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA
REFERENCE 3 (bases 1 to 160420)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-2001) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA
REFERENCE 4 (bases 1 to 160420)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-2002) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453, USA
COMMENT On May 17, 2002 this sequence version replaced gi:14522958.
FEATURES
source
BASE COUNT 48686 a 32379 c 32258 g 47097 t
ORIGIN
Query Match 9.4%; Score 316; DB 9; Length 160420;
Best Local Similarity 100.0%; Pred. No. 1.7e-172; Indels 0; Gaps 0;
Matches 316; Conservative 0; Mismatches 0;
QY 2 ACAGTGTTCGCGCTGCACCGCTCGGAGGCTGGGTGACCGCGCTAGAGTGAAGTACTTT 61
DB 5373 ACAGTGTTCGCGCTGCACCGCTCGGAGGCTGGGTGACCGCGCTAGAGTGAAGTACTTT 5314
QY 62 TTTATTGACAGCTGGCGCGATGCGCTTTAAAAACGCGAGGGCTCTATGCACCTCC 121
DB 5313 TTTATTGACAGCTGGCGCGATGCGCTTTAAAAACGCGAGGGCTCTATGCACCTCC 5254
QY 122 CTGGCGGTAGTTCTCCGACCTCAGCGGCTGGGTGCGCGCTCTCCAGGAGAGA 181
DB 5253 CTGGCGGTAGTTCTCCGACCTCAGCGGCTGGGTGCGCGCTCTCCAGGAGAGA 5194
QY 182 CAAACAGGTGTCCAGCTGGCAGCGCGCGCGCGCGCGCTCTCTGTATCCCGTAGCGC 241
DB 5193 CAAACAGGTGTCCAGCTGGCAGCGCGCGCGCGCGCTCTCTGTATCCCGTAGCGC 5134
QY 242 CCCTTGCGCGAGCGCGCGCGCGCGCGCTGTGTAGTAGAGCGCGCGCGCGCGCTGGT 301
DB 5133 CCCTTGCGCGAGCGCGCGCGCGCGCTGTGTAGTAGAGCGCGCGCGCGCGCTGGT 5074
QY 302 CGCCGCTCTCTCTCCG 317
DB 5073 CGCCGCTCTCTCTCCG 5058
RESULT 3
AL161942/c
LOCUS AL161942
DEFINITION Homo sapiens chromosome 20 clone RP11-83B5, *** SEQUENCING IN PROGRESS ***, 33 unordered pieces.
AL161942
ACCESSION AL161942
VERSION AL161942.5 GI:9931708
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.

```

SOURCE      human..
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       1 (bases 1 to 136357)
JOURNAL     Burton, J.
COMMENT     Direct Submission
            Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
            On Aug 28, 2000 this sequence version replaced gi:9863622.
            ----- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquerry@sanger.ac.uk
            ----- Project Information
            Center project name: BA8395
            ----- Summary Statistics
            Sequencing program: XGAP4; version 4.5
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 114775 bases at least Q40
            Consensus quality: 123431 bases at least Q30
            Consensus quality: 128697 bases at least Q20
            Insert size: 133157; sum-of-contigs
            Quality coverage: 2.20x in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 5425: contig of 5425 bp in length
            * 5426 5525: gap of 100 bp
            * 5526 9483: contig of 3958 bp in length
            * 9484 9583: gap of 100 bp
            * 9584 14280: contig of 4697 bp in length
            * 14281 14380: gap of 100 bp
            * 14381 17312: contig of 2932 bp in length
            * 17313 17412: gap of 100 bp
            * 17413 22848: contig of 5436 bp in length
            * 22849 22948: gap of 100 bp
            * 22949 26090: contig of 3142 bp in length
            * 26091 26190: gap of 100 bp
            * 26191 28215: contig of 2025 bp in length
            * 28216 28315: gap of 100 bp
            * 28316 33862: contig of 5547 bp in length
            * 33863 33962: gap of 100 bp
            * 33963 44145: contig of 10183 bp in length
            * 44146 44245: gap of 100 bp
            * 44246 53165: contig of 8920 bp in length
            * 53166 53265: gap of 100 bp
            * 53266 57412: contig of 4147 bp in length
            * 57413 57512: gap of 100 bp
            * 57513 60170: contig of 2658 bp in length
            * 60171 60270: gap of 100 bp
            * 60271 65985: contig of 5715 bp in length
            * 65986 66085: gap of 100 bp
            * 66086 69094: contig of 3009 bp in length
            * 69095 69194: gap of 100 bp
            * 69195 71736: contig of 2542 bp in length
            * 71737 71836: gap of 100 bp
            * 71837 74955: contig of 3119 bp in length
            * 74956 75055: gap of 100 bp
            * 75056 77325: contig of 2270 bp in length
            * 77326 77425: gap of 100 bp
            * 77426 80709: contig of 3284 bp in length
            * 80710 80809: gap of 100 bp
            * 80810 82943: contig of 2134 bp in length
            *
            * 82944 83043: gap of 100 bp
            * 83044 85218: contig of 2175 bp in length
            * 85219 85318: gap of 100 bp
            * 85319 89861: contig of 4543 bp in length
            * 89862 89961: gap of 100 bp
            * 89962 92824: contig of 2863 bp in length
            * 92825 92924: gap of 100 bp
            * 92925 95822: contig of 2898 bp in length
            * 95823 95922: gap of 100 bp
            * 95923 98130: contig of 2208 bp in length
            * 98131 98230: gap of 100 bp
            * 98231 100594: contig of 2364 bp in length
            * 100595 100694: gap of 100 bp
            * 100695 109043: contig of 8349 bp in length
            * 109044 109143: gap of 100 bp
            * 109144 112146: contig of 3003 bp in length
            * 112147 112246: gap of 100 bp
            * 112247 117183: contig of 4937 bp in length
            * 117184 117283: gap of 100 bp
            * 117284 121861: contig of 4578 bp in length
            * 121862 121961: gap of 100 bp
            * 121962 124258: contig of 2297 bp in length
            * 124259 124358: gap of 100 bp
            * 124359 127036: contig of 2678 bp in length
            * 127037 127136: gap of 100 bp
            * 127137 133357: contig of 6221 bp in length
            * 133358 133457: gap of 100 bp
            * 133458 136357: contig of 2900 bp in length.
            *
            FEATURES
            source
            1..136357
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="20"
            /clone="RP11-83B5"
            /clone.lib="RPC1-11.1"
            1..5425
            /note="assembly_fragment:01370"
            fragment_chain:1
            clone_end:SP6
            vector_side:left"
            5526..9483
            /note="assembly_fragment:01120"
            fragment_chain:1"
            9584..14280
            /note="assembly_fragment:00095"
            fragment_chain:1"
            14381..17312
            /note="assembly_fragment:01155"
            fragment_chain:1"
            17413..22848
            /note="assembly_fragment:00025"
            fragment_chain:2"
            22949..26090
            /note="assembly_fragment:00387"
            fragment_chain:2"
            26191..28215
            /note="assembly_fragment:00262"
            fragment_chain:3"
            28316..33862
            /note="assembly_fragment:00134"
            fragment_chain:3"
            33963..44145
            /note="assembly_fragment:00605"
            fragment_chain:4"
            44246..53165
            /note="assembly_fragment:00989"
            fragment_chain:4"
            53266..57412
            /note="assembly_fragment:00876"
            fragment_chain:5"
            57513..60170
            /note="assembly_fragment:00248"
            fragment_chain:5"

```

misc_feature 60271..65985
/note="assembly_fragment:01554
fragment_chain:6"
66086..69094
/note="assembly_fragment:00470
fragment_chain:6"
69195..71736
/note="assembly_fragment:01566
fragment_chain:7"
71837..74955
/note="assembly_fragment:00477
fragment_chain:7"
75056..77325
/note="assembly_fragment:00036"
77426..80709
/note="assembly_fragment:00390"
80810..82943
/note="assembly_fragment:00409"
83044..85218
/note="assembly_fragment:00437"
85319..89861
/note="assembly_fragment:00764"
89962..92824
/note="assembly_fragment:00782"
92925..95822
/note="assembly_fragment:00808"
95923..98130
/note="assembly_fragment:00831"
98231..100594
/note="assembly_fragment:00886"
100695..109043
/note="assembly_fragment:00992"
109144..112146
/note="assembly_fragment:01007"
112247..117183
/note="assembly_fragment:01102"
117284..121861
/note="assembly_fragment:01146"
121962..124258
/note="assembly_fragment:01320"
124359..127036
/note="assembly_fragment:01489"
127137..133357
/note="assembly_fragment:01498"
133458..136357
/note="assembly_fragment:01604"
BASE COUNT 38044 a 27571 c 27577 g 39953 t 3212 others
ORIGIN

Query Match 7.8%; Score 264; DB 2; Length 136357;
Best Local Similarity 100.0%; Pred. No. 5.2e-142; Indels 0; Gaps 0;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 ACAGTGTTCGGCGCTGCACCGCTCGGAGGCTGGGTGACCGCGTAGAAGTGAAGTACTTT 61
DB 9069 ACAGTGTTCGGCGCTGCACCGCTCGGAGGCTGGGTGACCGCGTAGAAGTGAAGTACTTT 9010
QY 62 TTTATTGACAGCTGGCGCGATGCCGCTTTAAAAACGCGAGGGGCTCTATGCACCTCC 121
DB 9009 TTTATTGACAGCTGGCGCGATGCCGCTTTAAAAACGCGAGGGGCTCTATGCACCTCC 8950
QY 122 CTGGCGGTAGTCTCTCCGACCTCAGCGGTCGGGTCTGCGCCCTCTCCAGGAGAGA 181
DB 8949 CTGGCGGTAGTCTCTCCGACCTCAGCGGTCGGGTCTGCGCCCTCTCCAGGAGAGA 8890
QY 182 CAACAGGTGTCCACGTGGCAGCGCGCCCGGGCGCCCTCTCTGTATCCGTAGCGC 241
DB 8889 CAACAGGTGTCCACGTGGCAGCGCGCCCGGGCGCCCTCTCTGTATCCGTAGCGC 8830
QY 242 CCCCTGGCCGACGCGCGCCCGG 265
DB 8829 CCCCTGGCCGACGCGCGCCCGG 8806

RESULT 4
G50805
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Contact: Michael Olivier, David R. Cox
Stanford Human Genome Center
Stanford University School of Medicine
4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
Tel: (650) 320-5800
Fax: (650) 320-5801
Email: olivier@shgc.stanford.edu
Primer A: TGCTTTAAATTCCTCCCTGCA
Primer B: AATGAGGCTCAGAGAGATGTGG
STS size: 348
PCR Profile:
Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700
Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Amplifrag Gold Polymerase: 0.07 units/ul
Total Vol: 5 ul
Buffer: MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

FEATURES
source
location/Qualifiers
1..496
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="10"
/clone_lib="Human"
132..479
132..154
primer_bind complement(457..479)
primer_bind 129 a 103 c 128 g 133 t 3 others
BASE COUNT 129 a 103 c 128 g 133 t 3 others
ORIGIN

Query Match 2.0%; Score 69; DB 11; Length 496;
Best Local Similarity 100.0%; Pred. No. 9.2e-28;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 630 GGAATTCCTTGGATTCAATTTTCAACCAAGGAGTGAAGCAAGATCAAGAGGAT 689
DB 21 GGAATTCCTTGGATTCAATTTTCAACCAAGGAGTGAAGCAAGATCAAGAGGAT 80
QY 690 GGTTCCTCA 698
DB 81 GGTTCCTCA 89

G50805 496 bp DNA linear STS 30-MAR-2000
SHGC-83858 Human Homo sapiens STS genomic, sequence tagged site.
G50805
G50805.1 GI:5221982
STS.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 496)
Olivier, M. and Cox, D.R.
Unpublished, Olivier, M., Cox, D.R. (2000)
Unpublished (2000)

Contact: Michael Olivier, David R. Cox
Stanford Human Genome Center
Stanford University School of Medicine
4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
Tel: (650) 320-5800
Fax: (650) 320-5801
Email: olivier@shgc.stanford.edu
Primer A: TGCTTTAAATTCCTCCCTGCA
Primer B: AATGAGGCTCAGAGAGATGTGG
STS size: 348
PCR Profile:
Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700
Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Amplifrag Gold Polymerase: 0.07 units/ul
Total Vol: 5 ul
Buffer: MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

BAC ends sequenced at TIGR from the RPC111 BAC library. Designed and developed at the Stanford Human Genome Center.


```

RESULT 5
AC125150
LOCUS      154748 bp      DNA      linear      HTG 26-JUN-2002
DEFINITION Mus musculus chromosome UNK clone RP24-35AK9, WORKING DRAFT
SEQUENCE, 8 unordered pieces.
ACCESSION AC125150
VERSION    AC125150.1 GI:21490651
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE     house mouse.
ORGANISM   Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1. (bases 1 to 154748)
McPherson,J.D. and Waterston,R.H.
TITLE      The sequence of Mus musculus clone
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 154748)
AUTHORS    McPherson,J.D. and Waterston,R.H.
JOURNAL    Direct Submission
TITLE      Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
AUTHORS    Parkway, St. Louis, MO 63108, USA
JOURNAL    3 (bases 1 to 154748)
REFERENCE  McPherson,J.D. and Waterston,R.H.
AUTHORS    Direct Submission
TITLE      Submitted (26-JUN-2002) Genome Sequencing Center, 4444 Forest Park
JOURNAL    Parkway, St. Louis, MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M_BB0354K09
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-terminator ET; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 151541 bases at least Q40
Consensus quality: 151970 bases at least Q30
Consensus quality: 152209 bases at least Q20
Insert size: 164000; agarose-fp
Insert size: 154048; sum-of-contigs
Quality coverage: 8.91 in Q20 bases; agarose-fp
Quality coverage: 8.71 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1201: contig of 1201 bp in length
* 1202: gap of unknown length
* 1302: contig of 2738 bp in length
* 4040: gap of unknown length
* 35917: contig of 31778 bp in length
* 35918: gap of unknown length
* 36017: contig of 4530 bp in length
* 40547: gap of unknown length
* 40548: gap of unknown length
* 51635: contig of 10988 bp in length
* 51636: gap of unknown length
* 51735: contig of 35396 bp in length
* 51736: gap of unknown length
* 87131: contig of 66133 bp in length
* 87132: gap of unknown length
* 87232: contig of 153364 bp in length
* 153365: gap of unknown length
* 153465: contig of 1284 bp in length.

```

```

FEATURES
source
Location/Qualifiers
1..154748
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-35AK9"
1..1201
/note="assembly_name:Contig10"
1302..4039
/note="assembly_name:Contig11"
clone_end:SP6
vector_side:right
4140..35917
/note="assembly_name:Contig14"
clone_end:T7
vector_side:right
36018..40547
/note="assembly_name:Contig12"
40648..51635
/note="assembly_name:Contig13"
51736..87131
/note="assembly_name:Contig15"
87232..153364
/note="assembly_name:Contig16"
153465..154748
/note="assembly_name:Contig18"
BASE COUNT 39535 a 38195 c 37106 g 39199 t 713 others
ORIGIN

Query Match 0.9%; Score 31; DB 2; Length 154748;
Best Local Similarity 100.0%; Pred. No. 2.4e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 614 GTTCCACTCCTCATCTGGAATTCCTCTGGA 644
|||||
DB 95867 GTTCCACTCCTCATCTGGAATTCCTCTGGA 95897

RESULT 6
F243081S03
LOCUS      1728 bp      DNA      linear      PRI 13-JUN-2001
DEFINITION Homo sapiens intrinsic factor-vitamin B12 receptor (CUBN) gene,
exons 5 and 6.
ACCESSION AF243083
VERSION    AF243083.1 GI:14388625
KEYWORDS   3 of 49
SEGMENT    Homo sapiens.
SOURCE     Homo sapiens.
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1728)
AUTHORS    Aminoff,M., Brady,S., Verroust,P.J., Moestrup,S.K. and Krahe,R.
TITLE      The genomic structure of the human CUBN gene encoding cubilin, the
intrinsic factor-vitamin B12 receptor
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 1728)
AUTHORS    Aminoff,M., Brady,S., Verroust,P.J., Moestrup,S.K. and Krahe,R.
TITLE      Direct Submission
JOURNAL    Submitted (08-MAR-2000) Division of Human Cancer Genetics, Ohio
State University, 420 West 12th Avenue, Columbus, OH 43210, USA

FEATURES
source
Location/Qualifiers
1..1728
/organism="Homo sapiens"
/db_xref="taxon:9606"
125..226
/gene="CUBN"
/number=5
/number=5
916..1019
/gene="CUBN"
/number=6
498 a 351 c 347 g 532 t
BASE COUNT 498 a 351 c 347 g 532 t
ORIGIN

```


misc_feature /note="encodes EGF repeat"
666..815
misc_feature /note="encodes EGF repeat"
816..944
misc_feature /note="encodes EGF repeat"
945..1076
misc_feature /note="encodes EGF repeat"
1077..1221
misc_feature /note="encodes EGF repeat"
1222..1322
misc_feature /note="encodes EGF repeat"
1323..1436
misc_feature /note="encodes EGF repeat"
1446..1793
misc_feature /note="encodes CUB domain"
1794..2147
misc_feature /note="encodes CUB domain"
2148..2474
misc_feature /note="encodes CUB domain"
2475..2819
misc_feature /note="encodes CUB domain"
2820..3167
misc_feature /note="encodes CUB domain"
3168..3518
misc_feature /note="encodes CUB domain"
3519..3857
misc_feature /note="encodes CUB domain"
3858..4196
misc_feature /note="encodes CUB domain"
4197..4553
misc_feature /note="encodes CUB domain"
4554..4883
misc_feature /note="encodes CUB domain"
4884..5237
misc_feature /note="encodes CUB domain"
5238..5582
misc_feature /note="encodes CUB domain"
5583..5957
misc_feature /note="encodes CUB domain"
5958..6299
misc_feature /note="encodes CUB domain"
6300..6674
misc_feature /note="encodes CUB domain"
6675..7031
misc_feature /note="encodes CUB domain"
7032..7379
misc_feature /note="encodes CUB domain"
7380..7733
misc_feature /note="encodes CUB domain"
7734..8090
misc_feature /note="encodes CUB domain"
8091..8438
misc_feature /note="encodes CUB domain"
8439..8783
misc_feature /note="encodes CUB domain"
8784..9134
misc_feature /note="encodes CUB domain"
9135..9494
misc_feature /note="encodes CUB domain"
9495..9857
misc_feature /note="encodes CUB domain"
9858..10208
misc_feature /note="encodes CUB domain"
10209..10556
misc_feature /note="encodes CUB domain"
10557..10895

Query Match 0.7%; Score 25; DB 9; Length 11344;
Best Local Similarity 100.0%; Pred. No. 0.069;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1236 CTGCCAGATGGAGGCACATGTGT 1260
|||||

Db 575 CTGCCAGATGGAGGCACATGTGT 599
RESULT 8
AC067747
LOCUS
DEFINITION Homo sapiens chromosome 10 clone RP11-406H21, complete sequence.
AC067747
VERSION AC067747.7 GI:20279381
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Smith, D.R.
Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
Unpublished
REFERENCE 2 (bases 1 to 147280)
AUTHORS Smith, D.R.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 3 (bases 1 to 147280)
AUTHORS Smith, D.R.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 4 (bases 1 to 147280)
AUTHORS Smith, D.R.
TITLE Direct Submission
JOURNAL Submitted (24-APR-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On Apr 24, 2002 this sequence version replaced gi:18767391.
FEATURES
source
Location/Qualifiers
1..147280
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-406H21"
/clone_lib="RPC1-11"
BASE COUNT 43059 a 27509 c 27812 g 48900 t
ORIGIN
Query Match 0.7%; Score 25; DB 9; Length 147280;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1236 CTGCCAGATGGAGGCACATGTGT 1260
|||||
Db 50565 CTGCCAGATGGAGGCACATGTGT 50589
RESULT 9
AC119647/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-137H14, *** SEQUENCING IN PROGRESS
***, 54 unordered pieces.
AC119647
VERSION AC119647.2 GI:21747172
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 152895)
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,

Carron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C. D., Cox, C., Coyle, M. D., Dathorne, S. R., David, R.,
 Davila, M. L., Davis, C., Davy-Carroll, L., Dederich, D. A.,
 Delaney, K. R., Delgado, O., Denna, A. L., Ding, Y., Dinh, H. H.,
 Douthwaite, K. J., Draper, H., Dugan-Rocha, S., Durbin, K. J.,
 Earhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
 Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mahoney, E., McLeod, M. P., Meador, M., Mel, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojubokan, I., Roife, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
 Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 152895)
 Worley, K. C.
 Direct Submission
 Submitted (30-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 152895)
 Worley, K. C.
 Direct Submission
 Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 14, 2002 this sequence version replaced gi:20340380.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GWPZ
 Center clone name: CH230-137H14
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembled program: Phrap; version 0.990329
 Consensus quality: 91714 bases at least Q40
 Consensus quality: 98583 bases at least Q30
 Consensus quality: 103878 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 54 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1041: contig of 1041 bp in length
 * 1042 1141: gap of unknown length
 * 1142 2262: contig of 1121 bp in length
 * 2263 2362: gap of unknown length
 * 2363 3497: contig of 1135 bp in length
 * 3498 3597: gap of unknown length
 * 3598 5210: contig of 1613 bp in length
 * 5211 5310: gap of unknown length
 * 5311 6936: contig of 1626 bp in length
 * 6937 7036: gap of unknown length
 * 7037 8291: contig of 1255 bp in length
 * 8292 8391: gap of unknown length
 * 8392 9663: contig of 1272 bp in length
 * 9664 9763: gap of unknown length
 * 9764 11369: contig of 1606 bp in length
 * 11370 11469: gap of unknown length
 * 11470 13073: contig of 1604 bp in length
 * 13074 13173: gap of unknown length
 * 13174 14521: contig of 1348 bp in length
 * 14522 14621: gap of unknown length
 * 14622 16274: contig of 1653 bp in length
 * 16275 16374: gap of unknown length
 * 16374 17405: contig of 1031 bp in length
 * 17406 17506: gap of unknown length
 * 17506 18547: contig of 1042 bp in length
 * 18548 18647: gap of unknown length
 * 18648 19808: contig of 1261 bp in length
 * 19809 20008: gap of unknown length
 * 20009 21460: contig of 1452 bp in length
 * 21461 21560: gap of unknown length
 * 21561 22822: contig of 1262 bp in length
 * 22823 22922: gap of unknown length
 * 22923 24377: contig of 1515 bp in length
 * 24378 24537: gap of unknown length
 * 24538 25832: contig of 1295 bp in length
 * 25833 25932: gap of unknown length
 * 25933 27104: contig of 1172 bp in length
 * 27105 27204: gap of unknown length
 * 27205 29052: contig of 1848 bp in length
 * 29053 30398: contig of 1246 bp in length
 * 30399 30498: gap of unknown length
 * 30499 32426: contig of 1928 bp in length
 * 32427 32526: gap of unknown length
 * 32527 34955: contig of 2429 bp in length
 * 34956 35055: gap of unknown length
 * 35056 36842: contig of 1587 bp in length
 * 36843 36742: gap of unknown length
 * 36743 37820: contig of 1078 bp in length
 * 37821 37920: gap of unknown length
 * 37921 39394: contig of 1474 bp in length
 * 39395 40928: contig of 1434 bp in length
 * 40929 41028: gap of unknown length
 * 41029 43380: contig of 2352 bp in length
 * 43381 43480: gap of unknown length
 * 43481 44748: contig of 1268 bp in length
 * 44749 44848: gap of unknown length
 * 44849 45927: contig of 2079 bp in length
 * 45928 47027: gap of unknown length
 * 47028 49481: contig of 2454 bp in length
 * 49482 49581: gap of unknown length
 * 49582 51923: contig of 2342 bp in length
 * 51924 52023: gap of unknown length
 * 52024 54126: contig of 2103 bp in length
 * 54127 54226: gap of unknown length
 * 54227 57214: contig of 2988 bp in length
 * 57215 57314: gap of unknown length
 * 57315 60024: contig of 2710 bp in length
 * 60025 60124: gap of unknown length

* 60125 63163: contig of 3039 bp in length
 * 63164 63263: gap of unknown length
 * 63264 66703: contig of 3440 bp in length
 * 66704 66803: gap of unknown length
 * 66804 68557: contig of 1754 bp in length
 * 68558 68657: gap of unknown length
 * 68658 72882: contig of 4225 bp in length
 * 72883 72982: gap of unknown length
 * 72983 76542: contig of 3560 bp in length
 * 76543 76642: gap of unknown length
 * 76643 80420: contig of 3778 bp in length
 * 80421 80520: gap of unknown length
 * 80521 84990: contig of 4470 bp in length
 * 84991 85090: gap of unknown length
 * 85091 88933: contig of 3843 bp in length
 * 88934 89033: gap of unknown length
 * 89034 96304: contig of 7271 bp in length
 * 96305 96404: gap of unknown length
 * 96405 101068: contig of 4664 bp in length
 * 101069 101168: gap of unknown length
 * 101169 105041: contig of 3873 bp in length
 * 105042 105141: gap of unknown length
 * 105142 110814: contig of 5673 bp in length
 * 110815 110914: gap of unknown length
 * 110915 115458: contig of 4544 bp in length
 * 115459 120911: contig of 5353 bp in length
 * 120912 121011: gap of unknown length
 * 121012 128307: contig of 7296 bp in length
 * 128308 128407: gap of unknown length
 * 128408 134219: contig of 5812 bp in length
 * 134220 134319: gap of unknown length
 * 134320 139473: contig of 5154 bp in length
 * 139474 139573: gap of unknown length
 * 139574 146170: contig of 6597 bp in length
 * 146171 146270: gap of unknown length
 * 146271 152895: contig of 6625 bp in length.

Query Match 0.7%; Score 25; DB 2; Length 152895;
 Best Local Similarity 100.0%; Pred. No. 0.078;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 614 GTTCCACTCTCATCTCGAATTC 638
 Db 122055 GTTCCACTCTCATCTCGAATTC 122031

RESULT 10
 HSNOTCH2
 LOCUS HSNOTCH2 622 bp DNA linear PRI 05-APR-1995
 DEFINITION H.sapiens NOTCH 2 gene.
 ACCESSION X80115
 VERSION X80115.1 GI:763114
 KEYWORDS notch2 gene.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Lardelli, M.T.
 Direct Submission
 TITLE Submitted (11-JUL-1994) M.T. Lardelli, Karolinska Institute,
 JOURNAL Developmental Biology Laboratory, CMB, 171 77 Stockholm, SWEDEN
 REFERENCE 2 (bases 1 to 622)
 Lardelli, M.T.
 AUTHORS The human NOTCH1, 2, and 3 genes are located at chromosome
 TITLE positions 9q34, 1p13-p11, and 19p13.2-p13.1 in regions of
 neoplasia-associated translocation
 JOURNAL Genomics 24 (2), 253-258 (1994)
 MEDLINE 95213014
 PUBMED 7698746
 FEATURES Location/Qualifiers
 source 1..622

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="p13-p11"
 /clone="B-11"
 /sub_clone="mc12"
 /haplotype="diploid"
 /tissue_type="placental"
 /clone_lib="human placental cosmid library in pWE15
 vector"
 185..317
 /gene="NOTCH2"
 185..317
 /gene="NOTCH2"
 /note="EGF repeat 31"
 167 a 122 c 132 g 201 t
 BASE COUNT
 ORIGIN
 Query Match 0.7%; Score 23; DB 9; Length 622;
 Best Local Similarity 100.0%; Pred. No. 0.89;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1231 CAGCCCTGCCAGATGAGGCAC 1253
 Db 242 CAGCCCTGCCAGATGAGGCAC 264
 RESULT 11
 HSNOTCH2
 LOCUS HSNOTCH2 789 bp mRNA linear PRI 21-NOV-1996
 DEFINITION Human Notch2 mRNA, partial cds.
 ACCESSION U77493
 VERSION U77493.1 GI:1679773
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Lemasson, I., Devaux, C. and Mesnard, J.M.
 TITLE Partial sequence of EGF-like repeat domain of human Notch2 mRNA
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 789)
 Lemasson, I., Devaux, C. and Mesnard, J.M.
 AUTHORS Direct Submission
 TITLE Submitted (05-NOV-1996) Laboratoire d'Immunologie des Infections
 JOURNAL Retrovirales, Institut de Biologie, 4, Boulevard Henri IV,
 Montpellier 34060 Cedex, France
 FEATURES Location/Qualifiers
 source 1..789
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="MT2"
 /tissue_type="T cell"
 <1..>789
 /gene="Notch2"
 <1..>789
 /gene="Notch2"
 /note="transmembrane protein; hn"
 /codon_start=-1
 /product="Notch2"
 /protein_id="AAB19224.1"
 /db_xref="GI:1679774"
 /translation="YVNSYTCCKAGFDGVCENNTNNECTESSCFNGGTCVDGINSFS
 CLCPVGTGTFCLHEINCESSHPCLNDGTCVDGLGYRCSPGLGYTKNCOTLVNLC
 RSPCKNGKTCVKQICLPSGWWGVCDFVNPSCDFAASRGVILVHLCOHSGVC
 INAGNTHYCCQPVGYTGYCYEQDLDECAINPCQHGATCSDFITGGYRCECVPGYQGVNC
 EYEDVDECONQPCQNGTCTDLVNHFKCSCPPGTRGLLCEINIDDC"
 180 a 192 c 214 g 203 t
 BASE COUNT
 ORIGIN
 Query Match 0.7%; Score 23; DB 9; Length 789;
 Best Local Similarity 100.0%; Pred. No. 0.9;

```

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1231 CAGCCCTGCCAGTAATGAGGCAC 1253
|||||
Db 682 CAGCCCTGCCAGTAATGAGGCAC 704

RESULT 12
HSA296168          924 bp  mRNA  linear  PRI 31-JAN-2001
LOCUS              Homo sapiens mRNA for keratin associated protein (KRTAP4.13 gene).
DEFINITION
ACCESSION          AJ296168
VERSION            1 GI:12655381
KEYWORDS            keratin associated protein; KRTAP4.13 gene.
SOURCE              human.
ORGANISM            Homo sapiens
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS            Rogers M.A., Langbein, L., Winter, H., Ehmann, C., Korn, B. and
                    Schweizer, J.
TITLE              Characterization of a cluster of human high/ ultrahigh keratin
                    associated proteins on chromosome 17q12-21
JOURNAL
REFERENCE          Rogers M.A.
AUTHORS            Rogers M.A.
TITLE              Direct Submission
JOURNAL            Submitted (16-OCT-2000) Rogers M.A., Research Program B, German
                    Cancer Research Center, Im Neuenheimer Feld 280, Heidelberg,
                    Germany 69120, GERMANY
COMMENT            Related genomic sequence: AC006070 (146419-146919nt).
FEATURES
                    Location/Qualifiers
                    1..1099
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="17"
                        /map="17q12-21"
                        /clone="sc123"
                        /tissue_type="scalp"
                        /gene="KRTAP4.4"
                        /protein_id="CAC27563.1"
                        /translation="MVNSCCGVCSDQGGLENCRRPSYCQTTCCRTTCRRPSCCRPS
                        CCRPQCQTCCTTCCHPSCCVSCCRPQCQTCCTTCRRPSCCVSCCRPQCQTCCTTCRRPSCCVSCCRP
                        GOSLCC"
                    polyA_signal 1058..1063
                    BASE COUNT 236 a 365 c 227 g 271 t
                    ORIGIN
Query Match          0.7%; Score 23; DB 9; Length 924;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1190 CCACCTGCTACAGGACCACTGC 1212
|||||
Db 558 CCACCTGCTACAGGACCACTGC 580

RESULT 14
AF308601          9722 bp  mRNA  linear  PRI 22-NOV-2000
LOCUS              Homo sapiens NOTCH 2 (N2) mRNA, complete cds.
DEFINITION
ACCESSION          AF308601 M99437
VERSION            1 GI:11275977
KEYWORDS
SOURCE              Homo sapiens.
ORGANISM            Homo sapiens
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS            Stifani, S., Blaumueller, C.M., Redhead, N.J., Hill, R.E. and
                    Artavanis-Tsakonas, S.
TITLE              Human homologs of a Drosophila Enhancer of split gene product
                    define a novel family of nuclear proteins
JOURNAL            Nat. Genet. 2 (2), 119-127 (1992)
MEDLINE            93265135
PUBMED             1303260
REFERENCE          2 (bases 1 to 9722)
AUTHORS            Blaumueller, C.M. and Mann, R.S.
TITLE              Complete Human Notch 2 (hN2) cDNA sequence
JOURNAL            Unpublished
REFERENCE          3 (bases 5440 to 7353)
AUTHORS            Blaumueller, C.M. and Artavanis-Tsakonas, S.

```

```
sig_peptide
misc_feature
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 04:05:09 ; Search time 464 Seconds
(without alignments)
16380.376 Million cell updates/sec

Title: US-09-930-020A-1

Perfect score: 3375

Sequence: 1 gacagtgtcgcggtgcac.....tgtccaccttgaaggtcttc 3375

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*

```

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3375	100.0	3375	24	ABK92207
2	1537	45.5	3485	23	AA81530
3	296	8.8	406	23	AA81529
4	32	0.9	1837	21	AA61803
5	32	0.9	1837	22	AAC99736
6	32	0.9	1837	22	ABL34888
7	25	0.7	1422	23	AA592463
8	25	0.7	11344	23	AA592464
9	23	0.7	394	21	AAC05374
					Prostate cancer-as
					DNA encoding novel
					DNA encoding novel
					DNA encoding mur
					Skin cell cDNA, SE
					Murine cDNA isolat
					DNA encoding novel
					DNA encoding novel
					Human secreted pro

10	23	0.7	6728	24	AA28942	Human MOL1b cDNA.
11	23	0.7	7410	24	AA28941	Human MOL1a cDNA.
12	22	0.7	1851	22	AA52135	Human polynucleoti
13	22	0.7	1851	22	AA53119	Human polynucleoti
14	20	0.6	498	21	AA59627	Human secreted pro
15	20	0.6	972	23	AA564374	DNA encoding novel
16	20	0.6	1503	24	ABU91801	Human lipase endot
17	20	0.6	1755	20	AA85838	cDNA encoding huma
18	20	0.6	2565	19	AAV41622	Nucleotide sequenc
19	20	0.6	2565	19	AA96940	cDNA encoding LIPG
20	20	0.6	2802	24	ABL39928	Human central cann
21	20	0.6	3193	22	AA159385	Human polynucleoti
22	20	0.6	4282	21	AA771113	Human ORFX ORF2668
23	20	0.6	4352	22	AA158667	Human polynucleoti
24	20	0.6	4377	22	AA160453	Human polynucleoti
25	20	0.6	14775	23	ABL17452	Drosophila melanog
26	20	0.6	39328	24	ABL91800	Human lipase endot
27	19	0.6	123	22	ABA74045	Human foetal liver
28	19	0.6	123	22	AAK22499	Human brain expres
29	19	0.6	123	22	AAK48667	Human bone marrow
30	19	0.6	123	22	AA154495	Probe #23181 used
31	19	0.6	123	24	ABS22406	Human genome-deriv
32	19	0.6	250	21	AA24350	Human secreted pro
33	19	0.6	489	23	ABL23947	Drosophila melanog
34	19	0.6	505	22	ABA61547	Human foetal liver
35	19	0.6	505	22	AAK09848	Human brain expres
36	19	0.6	505	22	AAK35742	Human bone marrow
37	19	0.6	505	22	AA141457	Probe #10143 used
38	19	0.6	505	24	ABS10004	Human genome-deriv
39	19	0.6	952	24	ABR35445	Human cDNA encodin
40	19	0.6	1272	20	AAK61296	Yeast morflp with m
41	19	0.6	1320	22	AAH53507	S. epidermidis Ope
42	19	0.6	1329	21	AA43184	Arabidopsis thalia
43	19	0.6	1329	21	AA43184	Arabidopsis thalia
44	19	0.6	1380	24	ABN92556	Staphylococcus epi
45	19	0.6	1509	23	AA589522	DNA encoding novel

ALIGNMENTS

RESULT 1
ABK92207
ID ABK92207 standard; DNA: 3375 BP.

XX AC ABK92207;

XX DT 15-AUG-2002 (first entry)

XX DE Prostate cancer-associated DNA sequence #93.

XX KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;
XX KW gene therapy; gene; ds.

XX OS Mammalia.

XX PN WO200230268-A2.

XX PD 18-APR-2002.

XX PF 12-OCT-2001; 2001WO-US32045.

XX PR 13-OCT-2000; 2000US-0687576.

XX PR 08-DEC-2000; 2000US-0733288.

XX PR 08-DEC-2000; 2000US-0733742.

XX PR 24-JAN-2001; 2001US-263957P.

XX PR 16-MAR-2001; 2001US-276791P.

XX PR 16-MAR-2001; 2001US-276888P.

XX PR 06-APR-2001; 2001US-281922P.

XX PR 24-APR-2001; 2001US-286214P.

XX PR 30-APR-2001; 2001US-0847046.

XX PR 04-MAY-2001; 2001US-288589P.

PA (EOSB-) EOS BIOTECHNOLOGY INC.
PI Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;
XX
XX WPI; 2002-471335/50.
XX P-PSDB; ABG61891.
XX
XX Detecting a prostate cancer-associated transcript in a cell in a
PT patient, useful for diagnosing prostate cancer (PC) or screening
PT modulators of PC, by determining if prostate cancer-associated genes
PT are expressed in a prostate tissue
XX
XX Claim 22; Page 376-377; 436pp; English.
PS
XX The present invention relates to methods of detecting a prostate
CC cancer-associated transcript in a cell from a patient. The method
CC comprises contacting a biological sample from the patient with
CC prostate cancer-associated polynucleotides (designated PC genes) that
CC selectively hybridize to a sequence that is at least 80% identical
CC to them. The prostate cancer-associated polynucleotide sequences
CC are differentially expressed in prostate tumour tissue or in
CC prostate cancer and are derived from the tissues of various
CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).
CC The methods of the invention are useful for diagnosing and treating
CC prostate cancer in mammals. The prostate cancer-associated genes are
CC useful for diagnosing or treating prostate cancer, as well as for
CC identifying modulators of prostate cancer or agents that inhibit
CC prostate cancer. The nucleic acid sequences are particularly useful
CC in gene therapy, as a vaccine or in antisense applications.
CC ABK92115-ABK92263 represent prostate cancer-associated polynucleotide
CC sequences.
XX
XX Sequence 3375 BP; 660 A; 944 C; 1062 G; 709 T; 0 other;
SQ

Query Match 100.0%; Score 3375; DB 24; Length 3375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACAGTGTTCGGGCTGACCGCTCGGAGCTGGGTGACCCCGTAGAAGTACTT 60
DB 1 GACAGTGTTCGGGCTGACCGCTCGGAGCTGGGTGACCCCGTAGAAGTACTT 60
QY 61 TTTTATTTCGACACCTGGGCGGATGCGCTTTTAAACACGAGGGGCTATGACACCTC 120
DB 61 TTTTATTTCGACACCTGGGCGGATGCGCTTTTAAACACGAGGGGCTATGACACCTC 120
QY 121 CTTGGCGGTAGTTCCTCCGACCTCAGCCGCGGTGCGGTGCGCGCCCTCTCCAGGAG 180
DB 121 CTTGGCGGTAGTTCCTCCGACCTCAGCCGCGGTGCGGTGCGCGCCCTCTCCAGGAG 180
QY 181 ACAACAGGTGTCACAGTGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
DB 181 ACAACAGGTGTCACAGTGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
QY 241 CCCCTTGGCCGAGCGCGCGCGCGGTCTGTGAGTAGAGCGCGCGCGCGCGCGCG 300
DB 241 CCCCTTGGCCGAGCGCGCGCGCGGTCTGTGAGTAGAGCGCGCGCGCGCGCGCG 300
QY 301 TCGCCGCTCTCCTTCGCTTATATCAACATGCCCCCTTCTGTGTGAGGCGGCTCTGT 360
DB 301 TCGCCGCTCTCCTTCGCTTATATCAACATGCCCCCTTCTGTGTGAGGCGGCTCTGT 360
QY 361 GTTTTCTGTTTCCAGAGTGGCCCATCTCTCCCTCTCCAGGAATCCATGTAACAAA 420
DB 361 GTTTTCTGTTTCCAGAGTGGCCCATCTCTCCCTCTCCAGGAATCCATGTAACAAA 420
QY 421 GAAACCATCGGAGAGTTTCAGCTGCCAGCAAAATGATGTGCTCGGCTGCAGTGGAC 480
DB 421 GAAACCATCGGAGAGTTTCAGCTGCCAGCAAAATGATGTGCTCGGCTGCAGTGGAC 480
QY 481 ATCATGTTTCTGTAGATGGTCTAACAGCGTCGGGAAAGGGAGCTTTGAAAGGTC 540
DB 481 ATCATGTTTCTGTAGATGGTCTAACAGCGTCGGGAAAGGGAGCTTTGAAAGGTC 540

QY 541 CACTTTGCCATCACAGTCTGTGACGGTCTGGACATCAGCCCCGAGAGGGTCAAGTGGGA 600
DB 541 CACTTTGCCATCACAGTCTGTGACGGTCTGGACATCAGCCCCGAGAGGGTCAAGTGGGA 600
QY 601 GCATTCACGTTCACTCCACTCTCATCTGGAATCCCTTCGATTCATTTCAACCCAA 660
DB 601 GCATTCACGTTCACTCCACTCTCATCTGGAATCCCTTCGATTCATTTCAACCCAA 660
QY 661 CAGGAAGTGAAGCAAGAAATCAAGAGGATGGTTTTCAAGAGGAGCGGCACGAGAGGAA 720
DB 661 CAGGAAGTGAAGCAAGAAATCAAGAGGATGGTTTTCAAGAGGAGCGGCACGAGAGGAA 720
QY 721 CTTGCTCTGAAATACCTTCTGCACAGAGGTTGCTTGGAGGCAAGAAATGCTTCTGTGCC 780
DB 721 CTTGCTCTGAAATACCTTCTGCACAGAGGTTGCTTGGAGGCAAGAAATGCTTCTGTGCC 780
QY 781 CAGATCCTCATCATCTGCTACTGATGGAAGTCCAGGGGATGTGCACTGCCATCCAAG 840
DB 781 CAGATCCTCATCATCTGCTACTGATGGAAGTCCAGGGGATGTGCACTGCCATCCAAG 840
QY 841 CAGCTGAAGGAAAGGGTGTCACTGTGTTGCTGTGGGGTCAAGTTTCCAGGTGGGAG 900
DB 841 CAGCTGAAGGAAAGGGTGTCACTGTGTTGCTGTGGGGTCAAGTTTCCAGGTGGGAG 900
QY 901 GAGCTCATGCTGCTGCCAGGAGCCTAGAGGCAACAGCTGCTGTGGTGAAGAGGTG 960
DB 901 GAGCTCATGCTGCTGCCAGGAGCCTAGAGGCAACAGCTGCTGTGGTGAAGAGGTG 960
QY 961 GAGGATGCCACCAACGGCTCTTTCAGCACCTTCAGCAGCTCGGCCCATCTGCTCCAGGCC 1020
DB 961 GAGGATGCCACCAACGGCTCTTTCAGCACCTTCAGCAGCTCGGCCCATCTGCTCCAGGCC 1020
QY 1021 ACGCCAGACTGCAGGCTGAGGCTCACCCCTGTGACACAGGACGCTGGAGATGTCCTCGG 1080
DB 1021 ACGCCAGACTGCAGGCTGAGGCTCACCCCTGTGACACAGGACGCTGGAGATGTCCTCGG 1080
QY 1081 GAGTTCGCTGGCAATGCCCATGCTGAGAGGATCGCGGGGACCCCTTCGGGTGCTGGCT 1140
DB 1081 GAGTTCGCTGGCAATGCCCATGCTGAGAGGATCGCGGGGACCCCTTCGGGTGCTGGCT 1140
QY 1141 GCACACTGTCCTTTCACAGCTGGAAGAGAGTGTCTTAACCCACCTGCACTGCTCTAC 1200
DB 1141 GCACACTGTCCTTTCACAGCTGGAAGAGAGTGTCTTAACCCACCTGCACTGCTCTAC 1200
QY 1201 AGGACCACTGCCAGGCGCTGTGACTCGAGCCCTGCGAGCCCTGCGAGGAGCATGTGT 1260
DB 1201 AGGACCACTGCCAGGCGCTGTGACTCGAGCCCTGCGAGCCCTGCGAGGAGCATGTGT 1260
QY 1261 CCAGAGGAGTGGACGGCTACCACTGCTTCCCTGCGCGCTGGCGCTGGAGGGAGGCTAAC 1320
DB 1261 CCAGAGGAGTGGACGGCTACCACTGCTTCCCTGCGCGCTGGCGCTGGAGGGAGGCTAAC 1320
QY 1321 TGTGCCCTGAAGCTGAGCCTGGAATCAGGCTCGAGCTCCTCTTCTGCTGACACACTCT 1380
DB 1321 TGTGCCCTGAAGCTGAGCCTGGAATCAGGCTCGAGCTCCTCTTCTGCTGACACACTCT 1380
QY 1381 GCGGGCACCACCTGTGACGGCTTCTCGCGGGCCAAAGTCTTGTGAAGCGGTTGTGCGG 1440
DB 1381 GCGGGCACCACCTGTGACGGCTTCTCGCGGGCCAAAGTCTTGTGAAGCGGTTGTGCGG 1440
QY 1441 GCGGTGCTGAGCGAGACTCTCGGGCCCGAGTGGGTGTGGCCACATACAGAGGGAGCTG 1500
DB 1441 GCGGTGCTGAGCGAGACTCTCGGGCCCGAGTGGGTGTGGCCACATACAGAGGGAGCTG 1500
QY 1501 CTGGTGGCGGTGCTGTGGGGAGTACCAGAGTGTGCTGACCTGCTGAGGAGCTCGAT 1560
DB 1501 CTGGTGGCGGTGCTGTGGGGAGTACCAGAGTGTGCTGACCTGCTGAGGAGCTCGAT 1560
QY 1561 GGCATTCCCTTCGTTGGTGGCCCCCACCCTGACGGGAGTGCCTTTCGGGAGCGGGCAGAG 1620
DB 1561 GGCATTCCCTTCGTTGGTGGCCCCCACCCTGACGGGAGTGCCTTTCGGGAGCGGGCAGAG 1620

Db	2701	GGCACTGAAATGGTGCTACCTTCTGGAATGCTGTGCCCGAGGTCTTGAATGTCTGC	2760
Qy	2761	TTCCCGCGGTGGCCAGGACCACTATTCTCACCTGAGGAGGAGGATGCCCAACTGCAGCC	2820
Db	2761	TTCCCGCGGTGGCCAGGACCACTATTCTCACCTGAGGAGGAGGATGCCCAACTGCAGCC	2820
Qy	2821	ATGCTGCTTAGACACAAGAACGAGCTGATGTCACCCACAAACGATGTTGTGAAAAATT	2880
Db	2821	ATGCTGCTTAGACACAAGAACGAGCTGATGTCACCCACAAACGATGTTGTGAAAAATT	2880
Qy	2881	TTGATGTGAAGTAATAACCCACTTTCGTACCTGCTGCTGCTGCTGCTGCTGCTGCTAT	2940
Db	2881	TTGATGTGAAGTAATAACCCACTTTCGTACCTGCTGCTGCTGCTGCTGCTGCTGCTAT	2940
Qy	2941	CTGCCACTTTCCCTTGAGGATAAACAGGGGCTCCTGAAGACTTAAATTTAGCGGCCCTGA	3000
Db	2941	CTGCCACTTTCCCTTGAGGATAAACAGGGGCTCCTGAAGACTTAAATTTAGCGGCCCTGA	3000
Qy	3001	CGTTCCTTTGCACACAATCAATGCTCGCCAGAAATGTTGTTGACACAGTAATAGCCCGAGAG	3060
Db	3001	CGTTCCTTTGCACACAATCAATGCTCGCCAGAAATGTTGTTGACACAGTAATAGCCCGAGAG	3060
Qy	3061	AGGCGCTTTACTAGAGCATCCTTTGGACGCGGAAGGCCACGGCCTTCAAGTTGCAAGACA	3120
Db	3061	AGGCGCTTTACTAGAGCATCCTTTGGACGCGGAAGGCCACGGCCTTCAAGTTGCAAGACA	3120
Qy	3121	GCACCTTTTCCACTTCCCAGAGACATTCTGGATGCATTTGCATTGAGCTCTGAAAGGGGG	3180
Db	3121	GCACCTTTTCCACTTCCCAGAGACATTCTGGATGCATTTGCATTGAGCTCTGAAAGGGGG	3180
Qy	3181	CTTGAGGAGCGTTTGTGACTTCTGGCGACTGCCCTTTTGTGTGGAAGAGACTTGGAAA	3240
Db	3181	CTTGAGGAGCGTTTGTGACTTCTGGCGACTGCCCTTTTGTGTGGAAGAGACTTGGAAA	3240
Qy	3241	GGTCTCAGACTGAATGTGACCAATTAACACAGCTTGGTTGATGATGGGGAGGGGCTGAGT	3300
Db	3241	GGTCTCAGACTGAATGTGACCAATTAACACAGCTTGGTTGATGATGGGGAGGGGCTGAGT	3300
Qy	3301	TGTCATGGGCCCGAGTCTGGAGGGCCAGTAAAAATCGTCTTGAGTCGTGAGCAGTGTCC	3360
Db	3301	TGTCATGGGCCCGAGTCTGGAGGGCCAGTAAAAATCGTCTTGAGTCGTGAGCAGTGTCC	3360
Qy	3361	ACCTTGAAGTCTTC 3375	
Db	3361	ACCTTGAAGTCTTC 3375	

RESULT 2
AAC81530

AA
AC
AAS81530:

DE DNA encoding novel human diagnostic protein #17334.

Homo sapiens

XX
PD
11-OCT-2001XX
PR 31-MAR-2000: 2000US-0540217.

PA (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG17343.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 1; SEQ ID No 17334; 103pp; English.
XX
CC the invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC the polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequencing data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3485 BP; 829 A; 850 C; 1030 G; 776 T; 0 other;

Query Match 45.5%; Score 1537; DB 23; Length 3485;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1537; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1215 AGGCCCTGTGACTCGCAGCCCTGCCAGAAATGGAGGCACATGTGTCCAGAAGGACTGGA 1274
DB 1949 AGGCCCTGTGACTCGCAGCCCTGCCAGAAATGGAGGCACATGTGTCCAGAAGGACTGGA 2008
QY 1275 CGGCTACCACTGCTCTGCCCGCTGGCCCTTTGGAGGGAGGCTAACTGTGCCCTGAAGCT 1334
DB 2009 CGGCTACCACTGCTCTGCCCGCTGGCCCTTTGGAGGGAGGCTAACTGTGCCCTGAAGCT 2068
QY 1335 GAGCCTGGAATGAGGCTGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1394
DB 2069 GAGCCTGGAATGAGGCTGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2128
QY 1395 GGACGGCTTCTCGGGGCCAAAGTCTTCGTGAAGCGTTTGTGGGCGCTGCTGAGCGA 1454
DB 2129 GGACGGCTTCTCGGGGCCAAAGTCTTCGTGAAGCGTTTGTGGGCGCTGCTGAGCGA 2188
QY 1455 GGACTCTCGGGCCGAGTGGGTGTGGCCACATACAGACGGAGCTGCTGTGGCGGTGCC 1514
DB 2189 GGACTCTCGGGCCGAGTGGGTGTGGCCACATACAGACGGAGCTGCTGTGGCGGTGCC 2248
QY 1515 TGTGGGGAGTACAGGATGTGCTTACCTGGTCTGGAGCCTGATGGCATTCCTTCCTCG 1574
DB 2249 TGTGGGGAGTACAGGATGTGCTTACCTGGTCTGGAGCCTGATGGCATTCCTTCCTCG 2308
QY 1575 TGTGGGGCCCACTGACGGGCAGTGCCTTGGCGGAGCGGCAGAGCGTGGCTTCGGGAG 1634
DB 2309 TGTGGGGCCCACTGACGGGCAGTGCCTTGGCGGAGCGGCAGAGCGTGGCTTCGGGAG 2368
QY 1635 CGCCACAGGACAGGCGGAGCGGCCACGTAGAGTGGTGGTTTTGCTCACTAGTCA 1694
DB 2369 CGCCACAGGACAGGCGGAGCGGCCACGTAGAGTGGTGGTTTTGCTCACTAGTCA 2428

QY 1695 CTCGAGGATGAGGTTGCGGGCCAGCGGCTCAGCAAGGGCGGAGAGCTGCTCTGCT 1754
DB 2429 CTCGAGGATGAGGTTGCGGGCCAGCGGCTCAGCAAGGGCGGAGAGCTGCTCTGCT 2488
QY 1755 GGGTGTAGGCAGTGAGGCCGTGCGGGCAGAGCTGGAGAGATCACAGGCAGCCCAAGCA 1814
DB 2489 GGGTGTAGGCAGTGAGGCCGTGCGGGCAGAGCTGGAGAGATCACAGGCAGCCCAAGCA 2548
QY 1815 TGTGATGGTCTACTCGSATCCTCAGGATCTGTTCACCAAAATCCCTGAGCTGAGGGGAA 1874
DB 2549 TGTGATGGTCTACTCGSATCCTCAGGATCTGTTCACCAAAATCCCTGAGCTGAGGGGAA 2608
QY 1875 GCTGTGAGCCGCGCAGGGCCAGGGTCCCGGACACAAAGCCCTGGACCTCTCTCATGTT 1934
DB 2609 GCTGTGAGCCGCGCAGGGCCAGGGTCCCGGACACAAAGCCCTGGACCTCTCTCATGTT 2668
QY 1935 GGACACTCTGCCCTCAGTAGGGCCCGAGAAATTTGCTCAGATCAGAGCTTTGTGAGAAG 1994
DB 2669 GGACACTCTGCCCTCAGTAGGGCCCGAGAAATTTGCTCAGATCAGAGCTTTGTGAGAAG 2728
QY 1995 CTGTGCCCTCCAGTTTGAAGTGAACCTGACGTGACACAGGTGGCTGGTGTGTATGG 2054
DB 2729 CTGTGCCCTCCAGTTTGAAGTGAACCTGACGTGACACAGGTGGCTGGTGTATGG 2788
QY 2055 CAGCCAGGTGCAGACTGCTTCGGGTGGACACCAACCCACCCGGGCTCGGATGCTGCG 2114
DB 2789 CAGCCAGGTGCAGACTGCTTCGGGTGGACACCAACCCACCCGGGCTCGGATGCTGCG 2848
QY 2115 GGCATTAGCCAGCCCTTACCTAGTGGGTGGGTGAGCGGCGACCCGCCCTGCTGCA 2174
DB 2849 GGCATTAGCCAGCCCTTACCTAGTGGGTGGGTGAGCGGCGACCCGCCCTGCTGCA 2908
QY 2175 CATCTATGACAAAGTGAATGACCGTCCAGAGGGGTGCCCGGCTGGTGTCCCCAAAGCTGT 2234
DB 2909 CATCTATGACAAAGTGAATGACCGTCCAGAGGGGTGCCCGGCTGGTGTCCCCAAAGCTGT 2968
QY 2235 GGTGTGCTCAGAGGGGAGAGCGGCGAGGATGAGCGGCTTCTGCCCGCAGAGCTGAG 2294
DB 2969 GGTGTGCTCAGAGGGGAGAGCGGCGAGGATGAGCGGCTTCTGCCCGCAGAGCTGAG 3028
QY 2295 GAACAAATGGCATCTCTGTCTTGGTCTGGGCGCTGTCTTAAGTGAGGCTCTGCG 2354
DB 3029 GAACAAATGGCATCTCTGTCTTGGTCTGGGCGCTGTCTTAAGTGAGGCTCTGCG 3088
QY 2355 GAGGCTTGCAGGTCCCGGGATTCCTGTATCCAGTGGGAGCTTACGCCACCTCGCGGTA 2414
DB 3089 GAGGCTTGCAGGTCCCGGGATTCCTGTATCCAGTGGGAGCTTACGCCACCTCGCGGTA 3148
QY 2415 CCACCAAGGAGCTCTCATTTAGTGGCTGTGTGAGAGCCCAAGCCAGCTCAACCTCTG 2474
DB 3149 CCACCAAGGAGCTCTCATTTAGTGGCTGTGTGAGAGCCCAAGCCAGCTCAACCTCTG 3208
QY 2475 CAACCAAGGAGCTCTCATTTAGTGGGAGCTGCTCTCAGAAATGGGAGCTACCGCTG 2534
DB 3209 CAACCAAGGAGCTCTCATTTAGTGGGAGCTGCTCTCAGAAATGGGAGCTACCGCTG 3268
QY 2535 CAAGTGTGCGGATGGCTGGGAGGGCCCCACTCGGAGAACCGGTGAGTGGAGCTTTGCTC 2594
DB 3269 CAAGTGTGCGGATGGCTGGGAGGGCCCCACTCGGAGAACCGGTGAGTGGAGCTTTGCTC 3328
QY 2595 TGTATGTGAGCCAGGATGATTTCTTGAAGCGCCCTGAGGCACATGCTCCCGTCA 2654
DB 3329 TGTATGTGAGCCAGGATGATTTCTTGAAGCGCCCTGAGGCACATGCTCCCGTCA 3388
QY 2655 GGAGGCGAGCAGCCGTACCCCTCCCAAGCAACTACAGAGAGGCTTGGGCACTGAAATGGT 2714
DB 3389 GGAGGCGAGCAGCCGTACCCCTCCCAAGCAACTACAGAGAGGCTTGGGCACTGAAATGGT 3448
QY 2715 GCCTACTTCTGGAATGTCTGTGCCCCAGGTCTTAG 2751
DB 3449 GCCTACTTCTGGAATGTCTGTGCCCCAGGTCTTAG 3485


```
Query Match      0.9%; Score 32; DB 21; Length 1837;
Best Local Similarity 100.0%; Pred. No. 9.1e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1146 CTGTCCTCTCTACAGCTGGAAGAGAGTGTTCC 1177
Db 942 CTGTCCTCTCTACAGCTGGAAGAGAGTGTTCC 973

RESULT 5
AAC99736
ID AAC99736 standard; cDNA; 1837 BP.
XX
AC AAC99736;
XX
DT 08-MAR-2001 (first entry)
XX
DE Skin cell cDNA, SEQ ID NO: 350.
XX
KW Mouse; skin cell; cytostatic; antiinflammatory; anti-HIV;
KW neotropic; neuroprotective; vulnerary; immunomodulatory; vaccine;
KW keratinocyte growth stimulation; cancer; angiogenesis inhibition;
KW inflammation; neurological disease; ss.
XX
OS Mus sp.
XX
PN WO200069884-A2.
XX
PD 23-NOV-2000.
XX
PF 15-MAY-2000; 2000WO-NZ00075.
XX
PR 14-MAY-1999; 99US-0312283.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Watson JD, Strachan L, Sleeman M, Kumble KD, Murison JG;
XX WPI; 2001-007495/01.
XX
DR New isolated polynucleotide used in the identification of genetic
PT disorders and encoding polypeptides used for treating inflammatory
PT disease, cancer and neurological diseases -
XX
PS Claim 1; Page 272-273; 352pp; English.
XX
CC The present polynucleotide encodes a polypeptide which is expressed in
CC mammalian skin cells. The polypeptide is useful for stimulating
CC keratinocyte growth and motility, inhibiting the growth of cancer cells,
CC modulating angiogenesis, inhibiting angiogenesis and vascularisation of
CC tumours, modulating skin inflammation, stimulating the growth of
CC epithelial cells, inhibiting the binding of human immunodeficiency virus
CC (HIV)-1 to leukocytes, and treating inflammatory disease, cancer and
CC neurological diseases. The polynucleotide can be used as a marker, in
CC the identification of genetic disorders, and for the design of
CC oligonucleotides for examining expression patterns.
XX
SQ Sequence 1837 BP; 370 A; 541 C; 557 G; 369 T; 0 other;

Query Match      0.9%; Score 32; DB 22; Length 1837;
Best Local Similarity 100.0%; Pred. No. 9.1e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1146 CTGTCCTCTCTACAGCTGGAAGAGAGTGTTCC 1177
Db 942 CTGTCCTCTCTACAGCTGGAAGAGAGTGTTCC 973

RESULT 6
ABL34888
ID ABL34888 standard; cDNA; 1837 BP.
XX
AC ABL34888;
XX
```

```
XX
DT 04-APR-2002 (first entry)
XX
DE Murine cDNA isolated from skin cells SEQ ID NO: 350.
XX
KW Human; rat; mouse; skin cell; skin wound; cancer; growth defect;
KW developmental defect; inflammatory disease; dermatological; vulnerary;
KW immunomodulator; anti-inflammatory; cytostatic; neuroprotective; gene;
KW ss.
XX
OS Mus sp.
XX
PN WO200190357-A1.
XX
PD 29-NOV-2001.
XX
PF 24-MAY-2001; 2001WO-NZ00099.
XX
PR 24-MAY-2000; 2000US-206650P.
PR 25-JUL-2000; 2000US-221232P.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Watson JD, Strachan L, Sleeman M, Onrust R, Murison JG, Kumble KD;
XX WPI; 2002-122020/16.
XX
DR New polynucleotides and polypeptides encoded by the polynucleotides
XX isolated from skin cells, useful for treating skin wounds, cancers,
XX growth and developmental defects, inflammatory diseases, or for
XX modulating immune responses -
XX
PS Claim 1; Page 227-228; 466pp; English.
XX
CC The present invention provides the protein and coding sequences of cDNAs
XX isolated from human, murine and rat skin cell libraries. The sequences
XX can be used in the development of therapeutic agents useful in the
XX treatment of skin diseases, including skin wounds, cancer, growth
XX defects, developmental defects and inflammatory diseases. The proteins
XX have important roles in the induction of hair growth, cell proliferation
XX and cell-cell interaction, in maintaining tissue integrity, in wound
XX healing and in modulating immune responses. The present sequence is a
XX cDNA of the invention.
XX
SQ Sequence 1837 BP; 370 A; 541 C; 557 G; 369 T; 0 other;

Query Match      0.9%; Score 32; DB 24; Length 1837;
Best Local Similarity 100.0%; Pred. No. 9.1e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1146 CTGTCCTCTCTACAGCTGGAAGAGAGTGTTCC 1177
Db 942 CTGTCCTCTCTACAGCTGGAAGAGAGTGTTCC 973

RESULT 7
AAS92463
ID AAS92463 standard; cDNA; 1422 BP.
XX
AC AAS92463;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #28267.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
```

```
XX 30-MAR-2001; 2001WO-US08631.
PF 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
PR 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
PR 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Dmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG28276.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 1; SEQ ID No 28267; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
PS polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS94197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Query Match 0.7%; Score 25; DB 23; Length 1422;
XX Best Local Similarity 100.0%; Pred. No. 0.21;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 1236 CTGCCAGAAATGGAGGCACATGTGTT 1260
Db 258 CTGCCAGAAATGGAGGCACATGTGTT 282

RESULT 8
AAS92464
ID AAS92464 standard; cDNA; 11344 BP.
XX
XX AAS92464;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #28268.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
```

```
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Dmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG28277.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 1; SEQ ID No 28268; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Query Match 0.7%; Score 25; DB 23; Length 11344;
XX Best Local Similarity 100.0%; Pred. No. 0.2;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 1236 CTGCCAGAAATGGAGGCACATGTGTT 1260
Db 575 CTGCCAGAAATGGAGGCACATGTGTT 599

RESULT 9
AAC05374
ID AAC05374 standard; cDNA; 394 BP.
XX
XX AAC05374;
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 9449.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
```



```
XX PA (GEST ) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX XX WPI; 2000-500381/45.
XX DR
XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX PS Claim 1; SEQ ID 9449; 71pp + CD-ROM; English.
XX CC
XX CC The present sequence is one of a large number of 5' ESTs derived from
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX CC identified within the present sequence. The 5' ESTs were prepared from
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences
XX CC derived from the 5' ends of mRNAs and even in those cases where longer
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX CC They are used to obtain upstream regulatory sequences and to design
XX CC expression and secretion vectors.
XX SQ
XX SQ Sequence 394 BP; 102 A; 84 C; 96 G; 112 T; 0 other;

Query Match          0.7%; Score 23; DB 21; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1231 CAGCCCTGCCAGAAATGGAGGCAC 1253
Db 282 CAGCCCTGCCAGAAATGGAGGCAC 304

RESULT 10
AAD28942
ID AAD28942 standard; cDNA; 6728 BP.
AC AAD28942;
XX 07-MAY-2002 (first entry)
XX Human MOL1b cDNA.
XX
XX Secreted molecule; MOL1b protein; MOLX; cardiomyopathy; atherosclerosis;
XX diabetes; chromosomal disorder; albinism; anaemia; psoriasis; scarring;
XX liver cirrhosis; neurodegeneration; osteoarthritis; organ rejection;
XX cerebral thrombosis; hypertension; systemic lupus erythematosus; asthma;
XX immune disease; ischaemia; immunodeficiency; rheumatoid arthritis; ulcer;
XX HIV; human immuno deficiency virus; hepatitis; haemostatic disease; pain;
XX haematopoietic; thrombolytic; thrombocytopaenia; Alzheimer's disease;
XX Parkinson's disease; Huntington's disease; muscular disease; stress;
XX ocular disease; growth disorder; depression; epilepsy; contraceptive;
XX vulnary; osteopathic; haemostatic; tranquiliser; antidepressant;
XX analgesic; vasotropic; hypotensive; gene therapy; chromosome 1; ss.
XX OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 5'UTR 1..12
FT /*tag= a
FT 13..4434
FT /*tag= b
FT /*tag= c
FT /*tag= d
FT sig_peptide
FT 13..87
FT /*tag= c
FT mat_peptide
FT 88..4431
FT /*tag= d

/product= "Human MOL1b mature protein"
4435..6728
/*tag= e

3'UTR
WO200206339-A2.
24-JAN-2002.
03-JUL-2001; 2001WO-US21249.
03-JUL-2000; 2000US-215854P.
03-JUL-2000; 2000US-215856P.
03-JUL-2000; 2000US-215902P.
07-JUL-2000; 2000US-216585P.
07-JUL-2000; 2000US-216586P.
07-JUL-2000; 2000US-216722P.
17-JUL-2000; 2000US-218622P.
17-JUL-2000; 2000US-218992P.
27-JUL-2000; 2000US-221285P.
14-FEB-2001; 2001US-268734P.
(CURA-) CURAGEN CORP.
Spaderna SK, Tchernov V, Liu X, Shenoy S, Spytek K, Zerhusen B;
Patturajan M, Taupier RJ, Rastelli L, Grosse WM, Szekeres ES;
Alsbrook J, Lepley DM, Shen L, Burgess CE, Shinkets RA;
Padigaru M;
WPI; 2002-155038/20.
P-PSDB; AAE18208.
Nucleic acids encoding secreted polypeptides, designated MOLX
polypeptides, useful for treating a MOLX-associated disorder, e.g.
cardiomyopathy, atherosclerosis, diabetes and metabolic disorders -
Claim 3; Page 14-15; 223pp; English.
The patent discloses nucleic acid sequences encoding novel secreted
molecule (MOL) polypeptides, designated MOLX polypeptides (i.e. a MOL
protein where X is an integer from 1 to 8). Sequences of the invention
are useful for treating or preventing a MOLX-associated disorder in
humans. They are useful for treating or preventing cardiomyopathy,
atherosclerosis and disorders related to cell signal processing and
metabolic pathway modulation. The MOLX antibodies are useful for
treating or preventing diabetes and disorders related to cell signal
processing and metabolic pathway modulation. MOLX sequences are useful
for the treatment or diagnosis of other MOLX-associated disorders, e.g.
chromosomal disorders, albinism, anaemia, liver cirrhosis, psoriasis,
scarring, neurodegeneration, osteoarthritis, organ rejection, cerebral
thrombosis, ischaemia, hypertension, systemic lupus erythematosus, immune
diseases, immunodeficiency, HIV (human immuno deficiency virus), viral,
bacterial and fungal infections, hepatitis, rheumatoid arthritis, asthma,
haematopoietic, thrombolytic, haemostatic diseases, thrombocytopaenia,
ulcers, Alzheimer's disease, Parkinson's disease, Huntington's disease,
ocular disease, muscular diseases, growth disorders, loss of libido,
stress, depression, pain and epilepsy. They are useful for preventing
chemotherapy side effects and as contraceptives. Sequences of the
invention are also useful for gene therapy. The present sequence
is a cDNA encoding human Notch-like protein, MOL1b. MOL1b gene is
localised on chromosome 1.
Sequence 6728 BP; 1571 A; 1647 C; 1668 G; 1842 T; 0 other;

Query Match          0.7%; Score 23; DB 24; Length 6728;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1231 CAGCCCTGCCAGAAATGGAGGCAC 1253
Db 3592 CAGCCCTGCCAGAAATGGAGGCAC 3614

RESULT 11
```


AAD28941

ID AAD28941 standard; cDNA; 7410 BP.

XX AC AAD28941;

XX DT 07-MAY-2002 (first entry)

XX DE Human MOLLA cDNA.

XX KW Secreted molecule; MOLLA protein; MOLX: cardiomyopathy; atherosclerosis;
 KW diabetes; chromosomal disorder; albinism; anaemia; psoriasis; scarring;
 KW liver cirrhosis; neurodegeneration; osteoarthritis; organ rejection;
 KW cerebral thrombosis; hypertension; systemic lupus erythematosus; asthma;
 KW immune disease; ischaemia; immunodeficiency; rheumatoid arthritis; ulcer;
 KW HIV; human immuno deficiency virus; hepatitis; haemostatic disease; pain;
 KW haematopoietic; thrombolytic; thrombocytopaenia; Alzheimer's disease;
 KW Parkinson's disease; Huntington's disease; muscular disease; stress;
 KW ocular disease; growth disorder; depression; epilepsy; contraceptive;
 KW vulnerability; osteopathic; haemostatic; tranquiliser; antidepressant;
 KW analgesic; vasotropic; hypotensive; gene therapy; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 1..7410

XX FT /tag= a

XX FT /product= "Human MOLLA protein"

XX FT sig_peptide 1..75

XX FT /tag= b

XX FT mat_peptide 76..7407

XX FT /tag= c

XX FT /product= "Human MOLLA mature protein"

XX FT WO200206339-A2.

XX PN 24-JAN-2002.

XX PD 03-JUL-2001; 2001WO-US21249.

XX PF 03-JUL-2000; 2000US-215854P.

XX PR 03-JUL-2000; 2000US-215856P.

XX PR 03-JUL-2000; 2000US-215902P.

XX PR 07-JUL-2000; 2000US-216585P.

XX PR 07-JUL-2000; 2000US-216586P.

XX PR 07-JUL-2000; 2000US-216722P.

XX PR 17-JUL-2000; 2000US-218622P.

XX PR 17-JUL-2000; 2000US-218992P.

XX PR 27-JUL-2000; 2000US-221285P.

XX PR 14-FEB-2001; 2001US-268734P.

XX PA (CURA-) CURAGEN CORP.

XX PI Spaderna SK, Tchernev V, Liu X, Shenoy S, Spytek K, Zerhusen B;

XX PI Patturajan M, Taupier RJ, Rastelli L, Grosse WM, Szekeres ES;

XX PI Alsobrook J, Lepley DM, Shen L, Burgess CE, Shinkets RA;

XX PI Padigaru M;

XX XX WPI: 2002-155038/20.

XX DR P-PSDB; AAE18207.

XX XX Nucleic acids encoding secreted polypeptides, designated MOLX

XX PT polypeptides, useful for treating a MOLX-associated disorder, e.g.

XX PT cardiomyopathy, atherosclerosis, diabetes and metabolic disorders

XX PS Claim 3; Page 9-11; 223pp; English.

XX XX The patent discloses nucleic acid sequences encoding novel secreted

XX CC molecule (MOL) polypeptides, designated MOLX polypeptides (i.e. a MOL

XX CC protein where x is an integer from 1 to 8). Sequences of the invention

XX CC are useful for treating or preventing a MOLX-associated disorder in

XX CC humans. They are useful for treating or preventing cardiomyopathy,

XX CC atherosclerosis and disorders related to cell signal processing and

XX CC metabolic pathway modulation. The MOLX antibodies are useful for

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC treating or preventing diabetes and disorders related to cell signal
 CC processing and metabolic pathway modulation. MOLX sequences are useful
 CC for the treatment or diagnosis of other MOLX-associated disorders, e.g.
 CC chromosomal disorders, albinism, anaemia, liver cirrhosis, psoriasis,
 CC scarring, neurodegeneration, osteoarthritis, organ rejection, cerebral
 CC thrombosis, ischaemia, hypertension, systemic lupus erythematosus, immune
 CC diseases, immunodeficiency, HIV (human immuno deficiency virus), viral,
 CC bacterial and fungal infections, hepatitis, rheumatoid arthritis, asthma,
 CC haematopoietic, thrombolytic, haemostatic diseases, thrombocytopaenia,
 CC ulcers, Alzheimer's disease, Parkinson's disease, Huntington's disease,
 CC ocular disease, muscular diseases, growth disorders, loss of libido,
 CC stress, depression, pain and epilepsy. They are useful for preventing
 CC chemotherapy side effects and as contraceptives. Sequences of the
 CC invention are also useful for gene therapy. The present sequence
 CC is a cDNA encoding human Notch-like protein, MOLLA.

XX SQ Sequence 7410 BP; 1691 A; 1989 C; 1983 G; 1747 T; 0 other;

Query Match 0.7%; Score 23; DB 24; Length 7410;

Best Local Similarity 100.0%; Pred. No. 1.9;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

..OY 1231 CAGCCCTGCCAGATGGAGGCAC 1253

DB 3577 CAGCCCTGCCAGATGGAGGCAC 3599

RESULT 12

AAK52135

ID AAK52135 standard; cDNA; 1851 BP.

XX AC AAK52135;

XX XX 06-NOV-2001 (first entry)

XX DT Human polynucleotide SEQ ID NO 680.

XX DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;

XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;

XX KW nervous system disorder; arthritis; inflammation; ss.

XX OS Homo sapiens.

XX PN WO200157190-A2.

XX XX 09-AUG-2001.

XX PD 05-FEB-2001; 2001WO-US04098.

XX PF 03-FEB-2000; 2000US-0496914.

XX XX 27-APR-2000; 2000US-0560875.

XX PR 20-JUN-2000; 2000US-0598075.

XX PR 19-JUL-2000; 2000US-0620325.

XX PR 01-SEP-2000; 2000US-0654936.

XX PR 15-SEP-2000; 2000US-0663561.

XX PR 20-OCT-2000; 2000US-0693325.

XX PR 30-NOV-2000; 2000US-0728422.

XX XX (HYSE-) HYSEQ INC.

XX PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

XX PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX XX WPI: 2001-476283/51.

XX DR P-PSDB; AAM79002.

XX XX Nucleic acids encoding polypeptides with cytokine-like activities,

XX PT useful in diagnosis and gene therapy

XX XX Claim 1; Page 2374-2376; 6221pp; English.

XX PS

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.

XX
 SQ Sequence 1851 BP; 379 A; 558 C; 574 G; 340 T; 0 other;

Query Match 0.7%; Score 22; DB 22; Length 1851;
 Best Local Similarity 100.0%; Pred. No. 5.7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1190 CCACCTGCTACAGGACCACCTG 1211
 |||||
 DB 1727 CCACCTGCTACAGGACCACCTG 1748

RESULT 13

AAK53119/c
 ID AAK53119 standard; cDNA; 1851 BP.

XX
 AC AAK53119;

XX
 DT 06-NOV-2001 (first entry)

XX
 DE Human polynucleotide SEQ ID NO 2648.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.

XX
 OS Homo sapiens.

XX
 PN WO200157190-A2.

XX
 PD 09-AUG-2001.

XX
 PF 05-FEB-2001; 2001WO-US04098.

XX
 PR 03-FEB-2000; 2000US-0496914.

XX
 PR 27-APR-2000; 2000US-0560875.

XX
 PR 20-JUN-2000; 2000US-0598075.

XX
 PR 19-JUL-2000; 2000US-0620325.

XX
 PR 01-SEP-2000; 2000US-0654936.

XX
 PR 15-SEP-2000; 2000US-0663561.

XX
 PR 20-OCT-2000; 2000US-0693325.

XX
 PR 30-NOV-2000; 2000US-0728422.

XX
 PA (HYSE-) HYSEQ INC.

XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

XX
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX
 DR WPI; 2001-476283/51.

XX
 DR P-PSDB; AAM79986.

XX Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -

XX
 PS Claim 1; Page 4903; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the

CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.

XX
 SQ Sequence 1851 BP; 340 A; 574 C; 558 G; 379 T; 0 other;

Query Match 0.7%; Score 22; DB 22; Length 1851;
 Best Local Similarity 100.0%; Pred. No. 5.7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1190 CCACCTGCTACAGGACCACCTG 1211
 |||||
 DB 125 CCACCTGCTACAGGACCACCTG 104

RESULT 14

AAK59627

ID AAK59627 standard; cDNA; 498 BP.

XX
 AC AAK59627;

XX
 DT 02-FEB-2001 (first entry)

XX
 DE Human secreted protein cDNA sequence #4.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; ss.

XX
 OS Homo sapiens.

XX
 PN WO2000056767-A1.

XX
 PD 28-SEP-2000.

XX
 PF 16-MAR-2000; 2000WO-US06828.

XX
 PR 19-MAR-1999; 99US-0125358.

XX
 PR 08-DEC-1999; 99US-0169616.

XX
 PA (HUMA-) HUMAN GENOME SCI INC.

XX
 PI Rosen CA, Ruben SM, Komatsoulis G;

XX
 DR WPI; 2000-602216/57.

XX
 DR P-PSDB; AAB34442.

XX Nucleic acid molecules encoding human secreted proteins, used in
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
 PT Parkinson's diseases and cancers -

XX
 PS Claim 1; Page 316; 384p; English.

XX The invention relates to the isolation of genes AAC59624-C59669 encoding
 CC XX human secreted proteins AAB34439-B34484. The genes can be used to
 CC generate fusion proteins by linking to the gene for the human
 CC immunoglobulin G Fc portion (AAC59615) for increasing the stability of
 CC the fusion protein as compared to the human protein only. The genes and
 CC proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated
 CC from a range of human tissues disclosed in the specification. The

CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.

XX SQ Sequence 498 BP; 109 A; 184 C; 93 G; 112 T; 0 other;

Query Match 0.6%; Score 20; DB 21; Length 498;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1829 CGGATCCTCAGGATCTGTC 1848
|||||
Db 360 CGGATCCTCAGGATCTGTC 379

RESULT 15

AAS64374/c
ID AAS64374 standard; cDNA; 972 BP.

XX AC AAS64374;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #178.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX P-PSDB; ABG00187.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity

XX PS Claim 1; SEQ ID No 178; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 972 BP; 252 A; 186 C; 263 G; 271 T; 0 other;

Query Match 0.6%; Score 20; DB 23; Length 972;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3049 AATGCCCCAGCAGAGGCCTTT 3068
|||||
Db 563 AATGCCCCAGCAGAGGCCTTT 544

Search completed: May 5, 2003, 05:07:01
Job time : 521 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 04:59:44 ; Search time 926 Seconds
(without alignments)
4306.664 Million cell updates/sec

Title: US-09-930-020A-1

Perfect score: 3375

Sequence: 1 gacagtgctgcgcgctgac.....tgccacctgaaggtcttc 3375

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 746064 seqs, 590810554 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications_NA:

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3375	100.0	3375	9 US-09-930-020A-1	Sequence 1, Appli
2	32	0.9	1837	9 US-10-152-661-350	Sequence 350, App
3	32	0.9	1837	9 US-09-866-050A-350	Sequence 350, App
4	20	0.6	531	9 US-09-918-995-19791	Sequence 19791, A
5	20	0.6	1120	9 US-09-938-842A-2842	Sequence 2842, Ap
6	20	0.6	3638	12 US-10-044-090-493	Sequence 493, App
7	19	0.6	123	10 US-09-864-761-29368	Sequence 29368, A
8	19	0.6	297	10 US-09-983-965-5125	Sequence 5125, Ap
9	19	0.6	505	10 US-09-864-761-12802	Sequence 12802, A
10	19	0.6	952	10 US-09-822-849A-583	Sequence 583, App
11	19	0.6	3047	10 US-09-818-143-7	Sequence 7, Appli
12	19	0.6	32367	9 US-10-158-160A-14	Sequence 14, Appl
13	19	0.6	172637	10 US-09-805-458A-3	Sequence 3, Appli
14	18	0.5	384	10 US-09-920-300A-1005	Sequence 1005, Ap
15	18	0.5	384	12 US-10-033-528-1005	Sequence 1005, Ap
16	18	0.5	401	9 US-09-946-807-1125	Sequence 1125, Ap
17	18	0.5	401	10 US-09-795-668-1125	Sequence 1125, Ap
18	18	0.5	401	10 US-09-795-686-1125	Sequence 1125, Ap
19	18	0.5	405	9 US-10-060-036-386	Sequence 386, App

20	18	0.5	408	9 US-09-918-995-17637	Sequence 17637, A
21	18	0.5	411	9 US-09-918-995-31656	Sequence 31656, A
22	18	0.5	439	9 US-09-918-995-27118	Sequence 27118, A
23	18	0.5	448	9 US-09-918-995-20362	Sequence 20362, A
24	18	0.5	455	9 US-10-060-036-191	Sequence 191, App
25	18	0.5	463	9 US-09-918-995-27657	Sequence 27657, A
26	18	0.5	469	9 US-09-918-995-10605	Sequence 10605, A
27	18	0.5	484	9 US-09-918-995-2281	Sequence 2281, Ap
28	18	0.5	491	9 US-09-918-995-1244	Sequence 1244, Ap
29	18	0.5	560	9 US-09-764-891-2501	Sequence 2501, Ap
30	18	0.5	600	10 US-09-864-761-8401	Sequence 8401, Ap
31	18	0.5	645	10 US-09-995-598-93	Sequence 93, Appl
32	18	0.5	717	10 US-09-815-242-7765	Sequence 7765, Ap
33	18	0.5	962	10 US-09-452-239-23	Sequence 23, Appl
34	18	0.5	1023	10 US-09-452-239-25	Sequence 25, Appl
35	18	0.5	1031	10 US-09-840-795-3	Sequence 3, Appl
36	18	0.5	1504	9 US-10-245-103-79	Sequence 79, Appl
37	18	0.5	1504	9 US-10-245-107-79	Sequence 79, Appl
38	18	0.5	1504	9 US-10-245-143-79	Sequence 79, Appl
39	18	0.5	1504	9 US-10-245-771-79	Sequence 79, Appl
40	18	0.5	1504	9 US-10-245-851-79	Sequence 79, Appl
41	18	0.5	1504	9 US-10-245-883-79	Sequence 79, Appl
42	18	0.5	1504	9 US-10-237-535-79	Sequence 79, Appl
43	18	0.5	1504	9 US-10-238-183-79	Sequence 79, Appl
44	18	0.5	1504	9 US-10-238-283-79	Sequence 79, Appl
45	18	0.5	1504	9 US-10-238-370-79	Sequence 79, Appl

ALIGNMENTS

RESULT 1

US-09-930-020A-1

; Sequence 1, Application US/09930020A

; Publication No. US20030077568A1

; GENERAL INFORMATION:

; APPLICANT: Gish, Kurt C.

; APPLICANT: Mack, David H.

; APPLICANT: Willson, Keith E.

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Colorectal Cancer. Compositions

; TITLE OF INVENTION: and Methods of Screening for Colorectal Cancer

; FILE REFERENCE: 018501-0031000US

; CURRENT APPLICATION NUMBER: US/09/930,020A

; CURRENT FILING DATE: 2001-08-14

; PRIOR APPLICATION NUMBER: US 09/663,733

; PRIOR FILING DATE: 2000-09-15

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 3375

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: CBF9

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (328)..(2751)

; OTHER INFORMATION: CBF9

; US-09-930-020A-1

Query Match 100.0%; Score 3375; DB 9; Length 3375;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACAGTGTTCGCGCTCACCGCTCGGAGCTGGGTGACCGCTAGAGTGAAGTACTT 60

|||||

1 GACAGTGTTCGCGCTCACCGCTCGGAGCTGGGTGACCGCTAGAGTGAAGTACTT 60

|||||

61 TTTTATTGACAGCTGGCGGCGATGGCGCTTTAAAAACGCGGGGCTCTATGCACCTC 120

|||||

61 TTTTATTGACAGCTGGCGGCGATGGCGCTTTAAAAACGCGGGGCTCTATGCACCTC 120

|||||

QY 121 CTTGGGGTAGTTCCTCCGACCTACGCGGGTCGGGTGGGTGCGCCCTCTCCAGAGAG 180
DB 121 CTTGGGGTAGTTCCTCCGACCTACGCGGGTCGGGTGGGTGCGCCCTCTCCAGAGAG 180
QY 181 ACAAACAGGTGTCACAGTGGACGCGCCCGCGGCGCCCTCTCTGATCCCTAGCG 240
DB 181 ACAAACAGGTGTCACAGTGGACGCGCCCGCGGCGCCCTCTCTGATCCCTAGCG 240
QY 241 CCCCTGGCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
DB 241 CCCCTGGCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
QY 301 TCCCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360
DB 301 TCCCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360
QY 361 GTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
DB 361 GTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
QY 421 GAAACCATCGGGAAGATTTTCAGTGCAGCAAAATGATGTGCTCGGCTGCAGTGGAC 480
DB 421 GAAACCATCGGGAAGATTTTCAGTGCAGCAAAATGATGTGCTCGGCTGCAGTGGAC 480
QY 481 ATCATGTTTCTGTAGATGGTCTACAGGTCGGGAAAGGAGCTTTGAAAGGTCCAAG 540
DB 481 ATCATGTTTCTGTAGATGGTCTACAGGTCGGGAAAGGAGCTTTGAAAGGTCCAAG 540
QY 541 CACTTTGCCATCACGTCTGACGCTGCGACATCAGCCCGGAGAGGTCAGAGTGGGA 600
DB 541 CACTTTGCCATCACGTCTGACGCTGCGACATCAGCCCGGAGAGGTCAGAGTGGGA 600
QY 601 GCATTCAGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTG 660
DB 601 GCATTCAGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTG 660
QY 661 CAGGAGTGAAGCAAGATCAGAGATGGTTTTCAGAGGAGGCGGAGAGAGAGAGAG 720
DB 661 CAGGAGTGAAGCAAGATCAGAGATGGTTTTCAGAGGAGGCGGAGAGAGAGAGAG 720
QY 721 CTTGCTCTGAAATPACCTTCTGCACAGAGGTTGCTTGGAGGCAAGATGTTCTGTGCC 780
DB 721 CTTGCTCTGAAATPACCTTCTGCACAGAGGTTGCTTGGAGGCAAGATGTTCTGTGCC 780
QY 781 CAGATCCTCATCATCTGCTACGTAGGGAAGTCCAGGGGATGTGCACTGCCATCCAAG 840
DB 781 CAGATCCTCATCATCTGCTACGTAGGGAAGTCCAGGGGATGTGCACTGCCATCCAAG 840
QY 841 CAGCTGAAGGAAAGGGGTGCTACTGTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG 900
DB 841 CAGCTGAAGGAAAGGGGTGCTACTGTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTG 900
QY 901 GAGCTGCATGCACTGGCCAGGAGCTAGAGGAGTACAGGAGTACAGGAGTACAGGAG 960
DB 901 GAGCTGCATGCACTGGCCAGGAGCTAGAGGAGTACAGGAGTACAGGAGTACAGGAG 960
QY 961 GAGGATGCCAACACCGGCTTTTCAGACCTCAGAGCTGCGCCATCTGCTCCAGCGCC 1020
DB 961 GAGGATGCCAACACCGGCTTTTCAGACCTCAGAGCTGCGCCATCTGCTCCAGCGCC 1020
QY 1021 ACCCCAGACTGAGGTCAGAGGCTACCCCTGTGACACAGGAGCTGGAGATGGTCCCG 1080
DB 1021 ACCCCAGACTGAGGTCAGAGGCTACCCCTGTGACACAGGAGCTGGAGATGGTCCCG 1080
QY 1081 GAGTTGCTGGCAATGCCCATGCTGGAGAGGATCGCGGAGCCCTTGGGTGCTGGGT 1140
DB 1081 GAGTTGCTGGCAATGCCCATGCTGGAGAGGATCGCGGAGCCCTTGGGTGCTGGGT 1140
QY 1141 GCACACTGCTCTTACAGCTGGAAGAGTGTCTTAAACCCACCTGCGACCTGCTAC 1200
DB 1141 GCACACTGCTCTTACAGCTGGAAGAGTGTCTTAAACCCACCTGCGACCTGCTAC 1200

QY 1201 AGGACCACCTGCCCCAGGCCCCCTGTGACTCGACGCCCTGCCAGANTGAGGACATGTGT 1260
DB 1201 AGGACCACCTGCCCCAGGCCCCCTGTGACTCGACGCCCTGCCAGANTGAGGACATGTGT 1260
QY 1261 CCAGAAGGACTGACAGGCTACCACTGCTGCTGCGGCTGCGCTTGGAGGAGGCTAAC 1320
DB 1261 CCAGAAGGACTGACAGGCTACCACTGCTGCTGCGGCTGCGCTTGGAGGAGGCTAAC 1320
QY 1321 TGTGCCCTGAAGCTGAGCCTGGAATGAGGCTGAGCTCCTCTCTCTCTCTCTCTCTCT 1380
DB 1321 TGTGCCCTGAAGCTGAGCCTGGAATGAGGCTGAGCTCCTCTCTCTCTCTCTCTCTCT 1380
QY 1381 GCGGGACCACTCTGACAGGCTTCTGCGGGCCAAAGTCTTCTGTAAGCGGTTTGTGCGG 1440
DB 1381 GCGGGACCACTCTGACAGGCTTCTGCGGGCCAAAGTCTTCTGTAAGCGGTTTGTGCGG 1440
QY 1441 GCGCTGCTGAGGAGGACTCTCGGGCCGAGTGGGTGGCCACATACAGAGGAGCTG 1500
DB 1441 GCGCTGCTGAGGAGGACTCTCGGGCCGAGTGGGTGGCCACATACAGAGGAGCTG 1500
QY 1501 CTGGTGGGCTGCTGCTGGGGAGTACCAGGATGTGCTGACCTGCTCTGGAGCTCGAT 1560
DB 1501 CTGGTGGGCTGCTGCTGGGGAGTACCAGGATGTGCTGACCTGCTCTGGAGCTCGAT 1560
QY 1561 GGCATTCCCTTCCGCTGGTGGCCCCACCTGACGGGCACTGCTTGGCGGAGCGGCGAG 1620
DB 1561 GGCATTCCCTTCCGCTGGTGGCCCCACCTGACGGGCACTGCTTGGCGGAGCGGCGAG 1620
QY 1621 CGTGGCTTCGGGAGCCACAGGACAGGACGCGGACGCTAGAGTGGTGGTGGTGG 1680
DB 1621 CGTGGCTTCGGGAGCCACAGGACAGGACGCGGACGCTAGAGTGGTGGTGGTGG 1680
QY 1681 CTCACCTGAGTCACTCCGAGGATGAGTGGGGGCGGCGGCTCAGCAAGGCGCGCA 1740
DB 1681 CTCACCTGAGTCACTCCGAGGATGAGTGGGGGCGGCGGCTCAGCAAGGCGCGCA 1740
QY 1741 GAGCTCTCTCTGCTGCTGAGGCTGAGGCTGCGGCGGAGCTGAGGAGGAGTACACA 1800
DB 1741 GAGCTCTCTCTGCTGCTGAGGCTGAGGCTGCGGCGGAGCTGAGGAGGAGTACACA 1800
QY 1801 GGAGCCCCAAGCATGTGATGCTTACTCGGATCTCTCAGGATCTGTTCAACCAATCCCT 1860
DB 1801 GGAGCCCCAAGCATGTGATGCTTACTCGGATCTCTCAGGATCTGTTCAACCAATCCCT 1860
QY 1861 GAGCTCAGGGGAGCTGTGACCGGCGGAGGCTGCGGAGCACAAGCCCTGGAC 1920
DB 1861 GAGCTCAGGGGAGCTGTGACCGGCGGAGGCTGCGGAGCACAAGCCCTGGAC 1920
QY 1921 CTCGCTCTCATGTTGACACCTCTGCTCAGTAGGCGCGGAGAAATTTGCTCAGATGCG 1980
DB 1921 CTCGCTCTCATGTTGACACCTCTGCTCAGTAGGCGCGGAGAAATTTGCTCAGATGCG 1980
QY 1981 AGCTTTGTGAGAGCTGTGCTTCCCTTCCAGTTTGGTGAACCTCAGCTGACACAGTCCGC 2040
DB 1981 AGCTTTGTGAGAGCTGTGCTTCCCTTCCAGTTTGGTGAACCTCAGCTGACACAGTCCGC 2040
QY 2041 CTGGTGGTGTATGGCAGGCTGACAGTGTGCTTGGGCTGACACCAACCCACCCCG 2100
DB 2041 CTGGTGGTGTATGGCAGGCTGACAGTGTGCTTGGGCTGACACCAACCCACCCCG 2100
QY 2101 GCTGCCATGCTGGGCGCATTAGCCAGGCCCCCTACTAGTGGGTGGGCTCAGCGCGC 2160
DB 2101 GCTGCCATGCTGGGCGCATTAGCCAGGCCCCCTACTAGTGGGTGGGCTCAGCGCGC 2160
QY 2161 ACCGCCCTGCTGCACATCTATGACAAAGTATGACCTCCAGAGGCTGCCGCCCTGGT 2220
DB 2161 ACCGCCCTGCTGCACATCTATGACAAAGTATGACCTCCAGAGGCTGCCGCCCTGGT 2220
QY 2221 GTCCCCAAAGCTGTGTGTGCTCAGGCGGAGAGGCTGACAGGATGACGCGCTTCT 2280
DB 2221 GTCCCCAAAGCTGTGTGTGCTCAGGCGGAGAGGCTGACAGGATGACGCGCTTCT 2280
QY 2281 GCCCAAGAGCTGAGGAACAATGTCATCTCTGTCTTGGTGGGCGCTGCGCTA 2340

Db 2281 GCGGAGAGCTGAGGAACAATGGCATCTCTGTCTGGTGGCGGTGGGCGCTGCTCTA 2340
QY 2341 ACTGAGGCTCTCGGAGGCTTCAGGTCCTCGGGATTCCTGATCCAGTCAGCTTAC 2400
Db 2341 AGTGAGGCTCTCGGAGGCTTCAGGTCCTCGGGATTCCTGATCCAGTCAGCTTAC 2400
QY 2401 GCGGAGCTCTCGGATCCAGGAGCTGCTCATTTAGTGGCTGTGTGGAGAGCCAGAG 2460
Db 2401 GCGGAGCTCTCGGATCCAGGAGCTGCTCATTTAGTGGCTGTGTGGAGAGCCAGAG 2460
QY 2461 CAGTCAACCTCTGCAACCCAGCCGCTGCATGAATGAGGCGAGCTGGCTCTGAGAA 2520
Db 2461 CAGTCAACCTCTGCAACCCAGCCGCTGCATGAATGAGGCGAGCTGGCTCTGAGAA 2520
QY 2521 GGGAGCTACCGCTGCAAGTGTGGGATGGCTGGGAGGCCCCCAGTGGAGAACCGTGCAG 2580
Db 2521 GGGAGCTACCGCTGCAAGTGTGGGATGGCTGGGAGGCCCCCAGTGGAGAACCGTGCAG 2580
QY 2581 TGGAGCTCTTGTCTGTATGTGTAGCCAGGAGTGTCTTGAGACCCCTTACAGAGCC 2640
Db 2581 TGGAGCTCTTGTCTGTATGTGTAGCCAGGAGTGTCTTGAGACCCCTTACAGAGCC 2640
QY 2641 ATGGCTCCGCTGAGGAGGAGGAGCCGCTACCCCTCCAGCACTACAGAGGCGCTG 2700
Db 2641 ATGGCTCCGCTGAGGAGGAGGAGCCGCTACCCCTCCAGCACTACAGAGGCGCTG 2700
QY 2701 GGCAGTGAATGGTGCCTTACCTTCTGGAATGTCTGTGCCAGGTCCTTAGAATGTCTGC 2760
Db 2701 GGCAGTGAATGGTGCCTTACCTTCTGGAATGTCTGTGCCAGGTCCTTAGAATGTCTGC 2760
QY 2761 TTCCGCGCGTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2820
Db 2761 TTCCGCGCGTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2820
QY 2821 ATGCTGCTTAGAGACAAAGAGCAGCTGATGTCACCCAGCAACAGATGTTGTGAAAAGTT 2880
Db 2821 ATGCTGCTTAGAGACAAAGAGCAGCTGATGTCACCCAGCAACAGATGTTGTGAAAAGTT 2880
QY 2881 TTGATGTGTAGTAAATACCCACTTCTGTACCTCTGTGCTGCTGCTGAGGCTATGTCTAT 2940
Db 2881 TTGATGTGTAGTAAATACCCACTTCTGTACCTCTGTGCTGCTGCTGAGGCTATGTCTAT 2940
QY 2941 CTGCCACCTTTCCCTTTGAGGAGTAAACAAGGGGCTCTGAAGACTTAAATTTAGCGGCC 3000
Db 2941 CTGCCACCTTTCCCTTTGAGGAGTAAACAAGGGGCTCTGAAGACTTAAATTTAGCGGCC 3000
QY 3001 CGTTCTTTGACACAAATCAATGCTGCCAGCAATGTTGTTGACACAGTAATGCCAGCAG 3060
Db 3001 CGTTCTTTGACACAAATCAATGCTGCCAGCAATGTTGTTGACACAGTAATGCCAGCAG 3060
QY 3061 AGGCTTTTACTAGACATCTTTTGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3120
Db 3061 AGGCTTTTACTAGACATCTTTTGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3120
QY 3121 GCAGCTTTTCCACTTCCCGAGAGATTTCTGATGATGATTTGATGATGATGATGATGATGAT 3180
Db 3121 GCAGCTTTTCCACTTCCCGAGAGATTTCTGATGATGATTTGATGATGATGATGATGATGAT 3180
QY 3181 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3240
Db 3181 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3240
QY 3241 GGTCTCAGACTGAATGTGACCAATTAACAGCTTGGTGTGATGATGATGATGATGATGATGAT 3300
Db 3241 GGTCTCAGACTGAATGTGACCAATTAACAGCTTGGTGTGATGATGATGATGATGATGATGAT 3300
QY 3301 TGTGATGGGCGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3360
Db 3301 TGTGATGGGCGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3360
QY 3361 ACCTTGAAGGCTTTC 3375
|||||

Db 3361 ACCTTGAAGGCTTTC 3375
RESULT 2
US-10-152-661-350
; Sequence 350, Application US/10152661
; Publication No. US20030022835A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c5
; CURRENT APPLICATION NUMBER: US/10/152.661
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 09/866.050
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/221.232
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: 60/206.650
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 09/312.283
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/NZ99/00051
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 09/188.930
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: 09/069.726
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 350
; LENGTH: 1837
; TYPE: DNA
; ORGANISM: Mouse
US-10-152-661-350
Query Match 0.9%; Score 32; DB 9; Length 1837;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1146 CTGTCCCTTCTACAGCTGGAAGAGAGTGTCC 1177
|||||
Db 942 CTGTCCCTTCTACAGCTGGAAGAGAGTGTCC 973
RESULT 3
US-09-866-050A-350
; Sequence 350, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866.050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 350
; LENGTH: 1837
; TYPE: DNA
; ORGANISM: Mouse
US-09-866-050A-350

Query Match 0.9%; Score 32; DB 9; Length 1837;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1146 CTGCTCTTCTACAGCTGGAAGAGAGTGTTC 1177
Db 942 CTGCTCTTCTACAGCTGGAAGAGAGTGTTC 973

RESULT 4

US-09-918-995-19791/c
; Sequence 19791, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19791
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(531)
; OTHER INFORMATION: n - A,T,C or G
US-09-918-995-19791

Query Match 0.6%; Score 20; DB 9; Length 531;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 752 TGCCTGGAGCAGCAATGCT 771
Db 114 TGCCTGGAGCAGCAATGCT 95

RESULT 5

US-09-938-842A-2842
; Sequence 2842, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRI1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2842
; LENGTH: 1120
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2842

Query Match 0.6%; Score 20; DB 9; Length 1120;
Best Local Similarity 100.0%; Pred. No. 4.3;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2865 ATGTTGTTGAAAAGTTTGA 2884
Db 394 ATGTTGTTGAAAAGTTTGA 413

RESULT 6

US-10-044-090-493
; Sequence 493, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 493
; LENGTH: 3638
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 400203.1
US-10-044-090-493

Query Match 0.6%; Score 20; DB 12; Length 3638;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2534 GCAAGTGTGGGATGCTGG 2553
Db 1203 GCAAGTGTGGGATGCTGG 1222

RESULT 7

US-09-864-761-29368/c
; Sequence 29368, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 29368
;; LENGTH: 123
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AP001537.1
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.96
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.9
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.98
;; OTHER INFORMATION: NT HIT: AL163218.2, EVALUE 2.00e-63
;; OTHER INFORMATION: EST_HUMAN HIT: A1807844.1, EVALUE 1.00e-01
;; OTHER INFORMATION: SWISSPROT HIT: P43403, EVALUE 7.10e+00
US-09-864-761-29368

Query Match 0.6%; Score 19; DB 10; Length 123;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2718 TACCTTCTGGAATCTGT 2736
|||||
Db 121 TACCTTCTGGAATCTGT 103

RESULT 8
US-09-983-965-5125
;; Sequence 5125, Application US/09983965
;; Patent No. US20020137160A1
;; GENERAL INFORMATION:
;; APPLICANT: Warren, Wesley C.
;; APPLICANT: Tao, Nengbing
;; APPLICANT: Byatt, John C.
;; APPLICANT: Mathalagan, Nagappan
;; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
;; FILE REFERENCE: 37-21(10297)C
;; CURRENT APPLICATION NUMBER: US/09/983,965
;; CURRENT FILING DATE: 2001-10-26
;; PRIOR APPLICATION NUMBER: US 09/465,231
;; PRIOR FILING DATE: 1999-12-15
;; PRIOR APPLICATION NUMBER: US 60/113,678
;; PRIOR FILING DATE: 1998-12-17
;; NUMBER OF SEQ ID NOS: 5912
;; SEQ ID NO 5125
;; LENGTH: 297
;; TYPE: DNA
;; ORGANISM: Bos taurus
;; FEATURE:
;; OTHER INFORMATION: Clone ID: 33-LIB34-046-Q1-E1-A2
US-09-983-965-5125

Query Match 0.6%; Score 19; DB 10; Length 297;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1488 CAGCAGGAGCTGCTGTG 1506
|||||
Db 38 CAGCAGGAGCTGCTGTG 56

RESULT 9
US-09-864-761-12802/c
;; Sequence 12802, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharron G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;; FILE REFERENCE: Aeomica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 12802
;; LENGTH: 505
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AP001537.1
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.96
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.9
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.98
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
US-09-864-761-12802

Query Match 0.6%; Score 19; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 15;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2718 TACCTTCTGAAGTCTGT 2736
|||||
Db 224 TACCTTCTGAAGTCTGT 206

RESULT 10

US-09-822-849A-583/c
; Sequence 583, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakr
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 583
; LENGTH: 952
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-583

Query Match 0.6%; Score 19; DB 10; Length 952;

Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 357 CTGTGTTTCCTGTTTCC 375
|||||
Db 173 CTGTGTTTCCTGTTTCC 155

RESULT 11

US-09-818-143-7/c
; Sequence 7, Application US/09818143
; Patent No. US20020019000A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Voikmuth, Wayne
; APPLICANT: Klingler, Tod M.
; TITLE OF INVENTION: POLYNUCLEOTIDES COEXPRESSED WITH MATRIX-REMODELING GENES
; FILE REFERENCE: PB-0004 CIP
; CURRENT APPLICATION NUMBER: US/09/818,143
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 3047
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1656953CB1
US-09-818-143-7

Query Match 0.6%; Score 19; DB 10; Length 3047;

Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 357 CTGTGTTTCCTGTTTCC 375
|||||
Db 2257 CTGTGTTTCCTGTTTCC 2239

RESULT 12

US-10-158-160A-14
; Sequence 14, Application US/10158160A
; Publication No. US20030059805A1
; GENERAL INFORMATION:
; APPLICANT: RAPPOLD-HOERBRAND, GUDRUN
; APPLICANT: RAO, ERCOLE
; TITLE OF INVENTION: HUMAN GROWTH GENE AND SHORT STATURE GENE REGION
; FILE REFERENCE: 108351-00004
; CURRENT APPLICATION NUMBER: US/10/158,160A
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 09/147,699
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: PCT/EP97/05355
; PRIOR FILING DATE: 1997-09-29
; PRIOR APPLICATION NUMBER: 60/027,633
; PRIOR FILING DATE: 1995-10-01
; PRIOR APPLICATION NUMBER: EP/97100583.0
; PRIOR FILING DATE: 1997-01-16
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 32367
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-158-160A-14

Query Match 0.6%; Score 19; DB 9; Length 32367;

Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 CCGAGCGCTGTCGCGCT 308
|||||
Db 12890 CCGAGCGCTGTCGCGCT 12908

RESULT 13

US-09-805-458A-3/c
; Sequence 3, Application US/09805458A
; Patent No. US20020042100A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al
; TITLE OF INVENTION: ISOLATED HUMAN ION CHANNEL PROTEINS,
; SEQUENCE 3, NUCLEIC ACID MOLECULES ENCODING HUMAN ION CHANNEL PROTEINS,
; AND USES THEREOF
; FILE REFERENCE: CLO00722
; CURRENT APPLICATION NUMBER: US/09/805,458A
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 172637
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(172637)
; OTHER INFORMATION: n = A,T,C or G
US-09-805-458A-3

Query Match 0.6%; Score 19; DB 10; Length 172637;

Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 754 CCTGGAGGCAGAAATGCTT 772
|||||
Db 122712 CCTGGAGGCAGAAATGCTT 122694

RESULT 14

US-09-920-300A-1005/c

; Sequence 1005, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1005
; LENGTH: 384
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-1005

Query Match 0.5%; Score 18; DB 10; Length 384;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 959 TGGAGGATGCCACCAACG 976
| | | | | | | | | | | | | | | | | | | | | |
Db 375 TGGAGGATGCCACCAACG 358

RESULT 15
US-10-033-528-1005/c
; Sequence 1005, Application US/10033528
; Patent No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033,528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1005
; LENGTH: 384
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-1005

Query Match 0.5%; Score 18; DB 12; Length 384;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 959 TGGAGGATGCCACCAACG 976
| | | | | | | | | | | | | | | | | | | | | |
Db 375 TGGAGGATGCCACCAACG 358

Search completed: May 5, 2003, 07:54:36
Job time : 1316 secs

THIS PAGE BLANK (USPTO)